DOCUMENT RESUME

ED 035 269 EM 007 066

AUTHOR	Andrade, Manuel; And Others
TITLE	Measurement of Listening Comprehension Via Television in Elementary School Spanish Instruction; Development of Programed Learning Materials for Use with Televised Spanish Instruction; Results in First, Second, and Third Year Research. Denver-Stanford Project on the Context of Instructional Television.
INSTITUTION	Denver County Public Schools, Colo.; Stanford Univ., Calif. Inst. for Communication Research.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	R-3; R-4; R-5; R-7; R-10
PUB DATE	Mar 64
NOTE	226p.
EDRS PRICE	EDRS Price MF-$1.00 HC-$11.40
DESCRIPTORS	Autoinstructional Aids, Conversational Language Courses, *Educational Programs, Experimental Programs, Instructional Materials, Instructional Television, *Language Instruction, Language Laboratories, Language Learning Levels, Language Proficiency, *Language Programs, Languages, Language Teachers, Language Tests, Listening Comprehension, Programed Instruction, Programs, *Spanish
IDENTIFIERS	Stanford Denver Project

ABSTRACT

The Stanford-Denver Project is a continuing study of elementary school Spanish instruction. These five documents report on various aspects of this study, especially the use of new technology to make instruction of Spanish more efficient. The first study details the construction of a listening comprehension test which could be given via television to students who can neither write nor read Spanish. The second discusses the development of programed learning materials for Spanish instruction and their effectiveness. The third document presents the results from an experiment testing different utilizations of instructional television. The fourth reports on experiments done to determine the proper time to begin study of written Spanish, and along with the fifth study, compares teacher instruction with automated instruction. A combination of both techniques is found to be preferable to either alone. The fifth study examines the use of audio aids in earlier grades and finds them efficacious. (BB)

ED035269

MEASUREMENT OF LISTENING COMPREHENSION VIA TELEVISION IN ELEMENTARY SCHOOL SPANISH INSTRUCTION

- MANUEL ANDRADE
- JOHN L. HAYMAN, JR.
- JAMES T. JOHNSON, JR.

EM007066

Denver Public Schools • Stanford University

RESEARCH ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

ED035269

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

MEASUREMENT OF LISTENING COMPREHENSION VIA TELEVISION IN ELEMENTARY SCHOOL SPANISH INSTRUCTION

by

Manuel Andrade
John L. Hayman, Jr.
James T. Johnson, Jr.

DENVER-STANFORD PROJECT ON THE CONTEXT OF INSTRUCTIONAL TELEVISION.

* * * * * *

School District Number One
City and County of Denver
Denver, Colorado

Institute for Communication Research
Stanford University
Stanford, California

Report Number 3
October 1961

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY William R. Spears, Denver Public Schools TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

SUMMARY

The Problem

Measurement is extremely important in the Denver-Stanford project since different teaching techniques are being compared. The audio-lingual methodology employed in teaching first year Spanish to fifth graders, however, creates special difficulties in test construction. Pupils can neither read nor write in Spanish, and translation to English is strictly avoided. The presence in classrooms of teachers of widely varying language backgrounds means that television administration of any pencil-and-paper test is necessary to assure similar testing conditions.

Results

Three pencil-and-paper, TV administered tests, designed to measure listening comprehension skills, were developed during the 1960-61 school year. Validity, reliability, comprehensiveness, discrimination, and certain other principles suggested by language experts were carefully considered during the development. The lack of an outside criterion against which to compare obtained results made assuring validity a problem. Careful definition of course objectives and a detailed content analysis showing the emphasis given each language element during TV instruction, however, provided for the strongest kind of construct validity. Extensive pre-testing and analysis of results of the actual TV administrations showed that the other criteria for test development and use were satisfied.

Two general types of test items were used. The first had pictures on each pupil's answer sheet, and the pupil responded by matching a picture to a statement spoken by the television instructor. The second type required choosing one of two possible alternatives (true or false, 1 or 2) in response to a statement or statements spoken by the instructor. In each test the picture items were substantially better than the others in terms of discrimination and the "1 or 2" alternative items were poorest.

MEASUREMENT OF LISTENING COMPREHENSION VIA TELEVISION IN ELEMENTARY SCHOOL SPANISH INSTRUCTION

by

Manuel Andrade
John L. Hayman, Jr.
James T. Johnson, Jr.

The Denver Public Schools and Stanford University's Institute for Communication Research are currently engaged in a joint research project on the context of instructional television. The purpose of the project is to learn how instructional television can best fit into the total teaching situation. A substantial amount of research has established that television is a very effective teaching medium. Ways of combining it with other educational activities must now be considered, and the Denver-Stanford project is a beginning effort in this direction. Kenneth E. Oberholtzer is principal investigator for the Denver Public Schools and Wilbur Schramm is principal investigator for Stanford University. This is one of a number of project progress reports.

1. THE PROBLEM

The Denver-Stanford project was wholly concerned during the first year with the beginning level (Level I) of Spanish instruction. More than 6,000 fifth grade pupils in the Denver Public Schools served as subjects.

Level I instruction in Denver is entirely audio-lingual; pupils hear and speak Spanish but do not read or write. Almost all conversation is in the second language. English is used occasionally by the instructor to explain but never to translate. The learning situation is kept as natural as possible; formal grammar, lists of vocabulary words, and isolated language elements out of context are shunned. This teaching methodology is based on experience gained by developers of the FLES (Foreign Languages in the Elementary School) program and is highly recommended (Andersson, Brady, et. al., 1954; Peloro, Hughes, Bowen, et. al., 1959; Brooks, 1960; Keesee, 1960).

Measurement is an unusual problem in this situation. Subjects cannot be asked to write any answers in Spanish since no writing skills have been developed. They cannot write in English since translation is avoided. They are unable to read in the second language. Furthermore, no hard and fast rules for development of language tests are available. Nelson Brooks states that "there has been no systematic synthesis, based on a careful examination of what language is and how it is learned, of a testing program that will measure success in second-language learning" (Brooks, 1960, p. 157). Finally, creating similar testing conditions for all 6,000 subjects is a challenge, since an unusual divergence in the classroom teachers' language backgrounds exists.

Circumstances, then, have dictated the development of a listening test requiring no reading, or writing, or translation; a test which can be given via television to assure uniform administration, pronunciation, visual presentation, and the like; and a test developed entirely by staff personnel without the benefit of prior experience or reported test development upon which to rely. In short, a locally developed listening comprehension test is dictated.

Is such a test suitable for measuring achievement at the beginning level of elementary school Spanish instruction? Experts in the field indicate that listening comprehension at this level is of extreme importance.

Brooks calls training in listening comprehension the first objective in language instruction and says that 50 per cent of the time in Level I should be devoted to developing listening skills (Brooks, 1960, p. 123). Huebener states that the learner's listening "with care and discrimination" is necessary to the development of good pronunciation and that both listening and pronunciation skills are "absolutely essential to accurate derivation of meaning from a second language and to later development of reading and writing skills" (Huebener, 1959, pp. 11-12).

Brooks suggests several principles which should be followed in building listening comprehension tests. Specifically, he says that this type of test should:

> contain no English, no dictation, no anecdotes, and no incorrect forms. The stimulus (should be) directed exclusively to the ear. The response (should be) the choice of a picture that correctly matches what is said (Brooks, 1960, p. 161).

His more general suggestions are:

> Test vocabulary only in a context of normal speech or by pairing naturally associated words and expressions in the foreign language.
>
> Present normal problems. To associate words with each other, to complete a form, a thought, or a description, to answer a reasonable question or give a reasonable rejoinder, to identify or recall forms that satisfy both sense and syntax, to write or say what is likely to be written or said -- such things can be called normal problems.
>
> Keep the test in the foreign language. Use English only for directions.
>
> Organize the test by putting together what goes together. Use only that material which is suited to the limits and purposes of the test and to the capacities of the group being measured.
>
> Give the student ample opportunity to show what he can do by preparing a sufficient number of questions that are different from each other and of the right degree of difficulty.
>
> Maintain a pattern of difficulty. Present the material in each part of the test so that the questions gradually increase in difficulty.
>
> Test one thing at a time. The responses we ask for are often very complicated; accuracy of measurement will be directly related to our success in holding all the variables constant save one and carefully noting its performance.
>
> Acquaint the student with all the techniques before the test. Examples should precede all questions requiring techniques with which the student may not be familiar (Brooks, 1960, pp. 166-167).

Validity must also be considered. Empirical validity in this situation is impossible since there is no outside criterion against which to compare obtained results. The best alternative is construct validity, which demands such careful definition of the required learning task that a skilled teacher can determine by inspection and pupil performance if a particular test item is appropriate and if the items collectively represent the total content taught.

To summarize, measuring instruments to evaluate instruction of fifth grade students in beginning Spanish were needed. Scarcity of reported work in this area necessitated development of these instruments entirely by the local staff. The audio-lingual teaching methodology and principles of foreign language evaluation determined these requirements for the instrument:

1. It should be a test of listening comprehension which could be administered via television and scored in the project office.

2. It should correspond closely to the subject matter taught. All objectives of the language instruction should be covered. The necessity for employing construct validity meant that careful and detailed definition of these objectives was essential.

3. The test should be in Spanish, with English used only for directions.

4. No reading or writing of Spanish should be employed.

5. All test items should be in the normal context of Spanish. Isolated vocabulary and language elements out of context should not be used.

6. The test should be reliable, and test items should discriminate.

Recognition of the importance of speaking skills should be mentioned, perhaps. Measurement of oral ability, an entirely different problem from that discussed herein, is necessary for complete evaluation and is receiving considerable attention in the Denver-Stanford Project. Work in this area will be described in a subsequent report.

2. CONSTRUCTION OF TESTS

a. Definition of Course Objectives

As mentioned before, the validity problem made the most careful and exact definition of course objectives necessary before test objectives could be established. The course objectives which were measureable through a paper-and-pencil listening comprehension test fell into two general groups: vocabulary acquisition and understanding of structure patterns.

Concerning vocabulary, Huebener says that the learning of a stock of words and expressions is essential to the proper functioning of linguistic skills (Huebener, 1959 p. 80). At the same time, experts in elementary level language instruction say that vocabulary items must not be taught in isolation nor by translation. "The learning of word lists and the pairing of two languages, item by item, must be strictly avoided. Vocabulary must be learned in context and not in lexical form" (Selvi, Briggs, et. al., 1958, p. 16).

Therefore, though translation and consideration of isolated language elements was strictly avoided in the Denver-Stanford project, pupils were expected to acquire a minimum vocabulary. This means understanding the meanings of individual words, with heavy emphasis on nouns and verbs at this level, and of clusters of words, including idioms, which convey special concepts when used together.

Determination of the specific vocabulary elements used and the emphasis given to each was necessary. For this purpose, the content of the television lessons was carefully analyzed. Each element was listed according to the lesson in which it first appeared and according to subsequent usage. Analysis of the first semester's content is shown in Appendix A.

The second general group of objectives, structure patterns, include such items as article and adjective agreement, verb endings, and word order. As with vocabulary, elements were taught in context rather than in isolation; no grammatical rules were taught as such but were emphasized through functional usage.

Agreement is probably the structure pattern most difficult for pupils to understand, since there is no comparable concept in English. "Agreement of adjectives and articles with their nouns is based on the fact that all nouns in Spanish are classified as masculine or feminine, a classification which often has no reference to the basic meaning of the noun but is an arbitrary arrangement" (MacRae, p. 18). Adjectives and articles must agree in number as well as gender, of course. Because early understanding of the agreement concept is important, it received considerable stress in the TV lessons.

Verb endings, which depend on tense, number, and person, are easier to understand because English usage is similar to Spanish. Spanish verbs, however, have many more endings than English verbs, and are thus more difficult. The present indicative in Spanish normally has six different endings, for example, whereas it normally has only two endings in English. In the TV series, only the first and third persons, singular and plural, in the present indicative were used.

Word order, or syntax, was not heavily stressed since the simple sentences taught in the beginning year of instruction used about the same word order as would English equivalents. Some emphasis, however, was given to the facts that Spanish adjectives often follow their referents and subject pronouns are often omitted since verb endings indicate the person acting.

This statement of course objectives, in general terms and in terms of language elements actually used, guided the development and use of test items. Its thoroughness made possible the strongest construct validity in test construction.

b. Construction of the First Semester Mid-term Test

Fifty items, divided into two 25-item, 15-minute telecasts, were selected for the first semester mid-term test. The telecasts were two days apart. A two part test was chosen partly through consideration of the attention span of the average fifth grader. Also, 50 items were considered a minimum for comprehensiveness, validity, and reliability. A further reason was to gain some of the advantages of the alternate-forms method of reliability estimation. According to Guilford, the reliability coefficient will indicate both internal consistency and stability of performance when similar test parts are separated by a relatively short time (but at least one day). The split-half method alone shows only internal consistency and overestimates because of the constancy of conditions encountered in a single test period (Guilford, 1954, pp. 373-378).

The following eight types of items were incorporated into the test:

1. Single picture items. Here, the pupil had a picture on his answer sheet. The instructor gave three statements, and the pupil chose the statement which correctly described the picture.

2. Multiple picture items. This type was similar to the first except that the pupil had three pictures on the answer sheet and the instructor gave one statement. The first two types of items were meant primarily to test extent of vocabulary and some verb usage.

3. Article agreement items. The instructor gave two sentences for each item. The sentences were alike except for the article, which was incorrect in one sentence. Pupils had a "1" and a "2" on the answer sheet for each item, and they responded by circling the number for the sentence they thought correct.

4. Items testing recognition of correct response. The instructor gave a sentence which required a response and a second person responded. If the response was correct, i.e., was a suitable answer to the sentence, the pupil

circled "sí" on his answer sheet. Otherwise he circled "no." These items were designed to test understanding of simple phrases and idioms such as, "Como esta usted?" An appropriate answer would be, "Muy bien, gracias," and an inappropriate answer would be, "De nada."

5. Simple identification items. A picture was shown on television, and the instructor identified it. If the identification was correct, the pupil circled "sí;" if it was incorrect, the pupil circled "no."

6. Adjective agreement items. The instructor gave a sentence which contained an adjective, and the pupil indicated by circling "sí" or "no" whether the adjective agreed with the noun it modified.

7. Word order items. The instructor gave two sentences which contained the same Spanish words. Word order was correct in one sentence and incorrect in the other. The pupil circled a "1" or a "2" to indicate whether he thought the first or second sentence was correct.

8. Verb usage items. Again two sentences were given, and they were alike except for the verb form. Only one verb form was correct. The pupil indicated his choice by circling a "1" or a "2."

One hundred and fifty-three test items were constructed and were tested first on 200 pupils who saw the televised Spanish series but were not in the research project. The biserial correlation coefficient of each item with the total score was computed as an index of item discrimination. Fifty items were then chosen as a proposed test, with discrimination, range of difficulty, complete coverage of subject material, and frequency distribution as the selection criteria. A mean of 35 correct answers (or 70 per cent) and a standard deviation of five were sought.

The 50 item form was then tested on 100 pupils as a final check. Again biserial correlations were computed. Five items were replaced because they failed to meet the selection criteria. The split-half reliability, with Spearman-Brown correction, of the test on this trial was .903, the average biserial correlation was + .585, and the average proportion of correct answers was .751.

The test was administered to the 6,000 pupils in the project in December, after approximately three months of instruction. The questions and test form are in Appendix B . In the actual administration, the split-half reliability with Spearman-Brown correction was .834, the average biserial correlation was + .439, and the average proportion of correct answers was .630. The standard deviation was 7.673.

In checking all reliabilities, the 50 item tests were randomly divided into two 25-item parts. The parts were equivalent, and approximately half of each was administered on separate days. The test was slightly speeded since pupils were required to respond to each item within a specified time. Gulliksen has given this formula to adjust the reliability coefficient for slight speeding:

$$R_m = R_{xx} - \frac{M_u}{\sigma^2_x}$$

where R_{xx} = the split-half reliability with Spearman-Brown correction and M_u = the mean number of items unattempted at the end of the test. If all items are completed, $R_m = R_{xx}$. The reliabilities reported here for the tests developed were therefore not affected by speeding since all items were completed (Guilford, 1954, pp. 391-392).

Table 1 gives a comparison between the final pre-test and the actual administration of biserial correlations and proportions who passed for each item.

Sections A and B of the test were given during the first 15-minute telecast and Sections C, D, E, F, and G were given the second telecast. Though the test was generally satisfactory, the need for several improvements became apparent.

Items were numbered by section, as listed in Table 1. This caused some confusion since several items had the same number. The obvious solution was to number from one to fifty.

Items with "1" and "2" as a possible response also proved somewhat confusing during administration via TV, though they had seemed satisfactory in the face-to-face pre-testing.

Part two of the test had five sections, with the result that as much time was required for instructions and examples as for administration of actual items. Since strict timing was required to give 25 items in 15 minutes, greater efficiency would result if the same comprehensiveness could be attained with fewer sections.

Interviews with a few of the pupils after the test showed that they found the items with pictures more interesting, and a check on the biserial correlations revealed that this type of item discriminated better than the others. The items with pictures on the test form (Sections A and C) had an average biserial correlation of .492, whereas the other items had an average of .372. It was determined that verb usage, article agreement, adjective agreement, and word order, in addition to simple identification, could be tested with the picture-type item. Elimination of several different item types and increased discrimination seemed possible through use of more picture items.

c. Construction of the First Semester Final

The first semester final tested content from the complete semester, i.e., it included the subject matter covered in the first test plus that covered in the second part of the semester. The 24 best discriminating items from the mid-term were included, and 26 new items were constructed.

Based on the results of the first test, the number of picture items was

TABLE 1.

BISERIAL CORRELATIONS OF AND PROPORTION WHO PASSED EACH ITEM ON THE FINAL PRE-TEST AND THE ACTUAL ADMINISTRATION -- FIRST SEMESTER MID-TERM

Item	Biserial Correlation Coefficient		Proportion Who Passed Item	
	Pre-Test	Final Form	Pre-Test	Final Form
A 1	.803	.539	.865	.535
2	.650	.349	.852	.705
3	.064	.505	.963	.790
4	.660	.559	.852	.875
5	.578	.470	.951	.740
6	.582	.334	.729	.580
7	.484	.224	.803	.790
8	.331	.331	.840	.750
9	.517	.499	.852	.670
10	.587	.526	.618	.510
11	[illegible]	.605	.790	.760
12	[illegible]	.381	.552	.495
13	.629	.714	.605	.650
[illegible]	.728	.745	.580	.805
[illegible]	[illegible]	.341	.448	.460
16	[illegible]	.692	.469	.550
B 1	.910	.054	.926	.545
2	.681	.404	.889	.450
3	.821	.223	.889	.465
4	.805	.430	.803	.495
5	.795	.571	.889	.610
6	.470	.005	.827	.730
7	.708	.413	.914	.415
8	.506	.459	.704	.820

Item	Biserial Correlation Coefficient		Proportion Who Passed Item	
	Pre-Test	Final Form	Pre-Test	Final Form
C 1	.390	.672	.840	.690
2	.371	.677	.926	.810
3	.888	.423	.877	.445
4	.431	.492	.865	.495
5	.849	.600	.790	.615
6	.773	.381	.753	.745
7	.446	.729	.840	.840
8	.839	.296	.815	.470
9	.518	.552	.630	.505
10	.716	.295	.827	.515
11	.767	.439	.753	.530
12	.528	.394	.445	.410
D 1	.739	.114	.914	.720
2	.694	.203	.840	.685
3	.396	.199	.679	.645
4	.120	.256	.556	.460
5	.762	.515	.810	.375
E 1	.992	.642	.840	.410
2	.372	.400	.692	.870
3	.482	.581	.690	.875
4	.762	.760	.810	.905
F 1	.613	.520	.655	.765
2	.291	.438	.580	.745
G 1	.449	.326	.543	.520
2	.481	.367	.605	.720
3	.545	.317	.296	.515

increased. The new items were chosen in much the same way as before. Seventy-eight items were pre-tested on 125 pupils, and 26 items were selected on the basis of discrimination, range of difficulty, coverage of subject material, and frequency distribution. Again a mean of 35 correct answers and a standard deviation of five were sought. The average pre-test biserial correlation for items selected was + .542.

This test was administered via television in two parts on the last two days of the semester. In the actual administration, the average biserial correlation was + .540, and the split-half reliability coefficient, with Spearman-Brown correction, was .899. The average proportion of correct answers was .720.

A comparison of pre-test and final administration biserial correlations and passing proportions for each item is given in Table 2. Sections A and B were administered during the first telecast, and Sections C, D, and E were administered during the second. Sections A, B, and C consisted of items with pictures on the test form, Section D consisted of type (6) items, and Section E consisted of type (5) items. The test questions and answer form are in Appendix C.

Again the picture items showed better discrimination. The average biserial correlation coefficient for picture items was + .570, while that of non-picture items was + .471.

The Pearson product moment correlation coefficient between the first semester mid-term and final examinations was .735.

d. Construction of the Second Semester Final

Only one test was given during the second semester -- a final exam during the third week in May, and it was constructed in the same way as the first semester exams.

Two hundred and thirty items were first prepared and were tried on 115

TABLE 2.

BISERIAL CORRELATIONS OF AND PROPORTION WHO PASSED EACH ITEM ON THE FINAL PRE-TEST AND THE ACTUAL ADMINISTRATION -- FIRST SEMESTER FINAL

Item	Biserial Correlation Coefficient		Proportion Who Passed Item	
	Pre-Test	Final Form	Pre-Test	Final Form
A 1	.815	.332	.936	.867
2	.558	.460	.928	.874
3*	.559	.330	.875	.927
4*	.745	.669	.805	.800
5*	.505	.523	.790	.834
6*	.605	.748	.760	.707
7*	.470	.649	.740	.780
8*	.499	.542	.670	.747
9*	.714	.846	.650	.647
10*	.692	.702	.550	.667
11*	.526	.621	.510	.507
12*	.539	.652	.535	.547
B13	.706	.418	.960	.900
14	.778	.563	.936	.907
15	.630	.641	.904	.780
16	.680	.573	.896	.747
17*	.729	.668	.840	.900
18*	.677	.341	.810	.827
19*	.381	.623	.745	.734
20*	.672	.534	.690	.840
21*	.600	.628	.615	.720
22*	.439	.713	.530	.574
23*	.552	.451	.505	.640
24*	.423	.459	.445	.760
25*	.394	.537	.410	.474
C 26	.612	.797	.929	.827
27	.482	.761	.766	.727
28	.525	.552	.631	.674
29	.536	.577	.603	.654
30	.528	.340	.589	.507
31	.475	.225	.582	.667
32	.447	.602	.567	.600
33	.400	.717	.532	.500
34	.401	.623	.468	.654
35	.304	.527	.404	.600
D 36	.312	.322	.915	.894
37	.576	.608	.872	.734
38	.585	.305	.837	.740
39	.394	.557	.730	.580
40	.523	.477	.723	.667
41	.465	.561	.681	.494
42	.283	.368	.610	.700
43*	.571	.551	.610	.840
44*	.413	.551	.415	.574
E 45*	.760	.646	.905	.914
46*	.581	.329	.875	.880
47	.365	.298	.858	.854
48	.376	.540	.808	.667
49	.356	.616	.645	.780
50*	.642	.432	.410	.587

*Items used in the first semester mid-term.

pupils. Fifty items were then selected from these with the same criteria as before, and the fifty item form was tried on 100 pupils. Several items were changed as a result of the second trial. Again, the number of picture items was increased.

In the final pre-test of this exam, the average biserial correlation was .526 and the average proportion of correct answers was .698. In the actual administration of the test, the average biserial correlation was .520 and the average proportion of correct answers was .694. The split-half reliability, with Spearman-Brown correction, was .881. A comparison of biserial correlations and proportion of correct answers is given in Table 3.

Sections A, B, and C were given during the first 15-minute telecast, and Sections D, E, and F during the second telecast. Sections A, B, D, and E consisted of items with pictures on the test form. Section C consisted of type (8) items and Section F consisted of type (6) items. The test form and questions are in Appendix D.

On this test the difference between picture and non-picture items was even more dramatic than before. The average biserial correlation for picture items was .570 and that for non-picture items only .295.

Type (8) items test verb agreement by having the pupil circle a "1" or a "2" to indicate which of two sentences is correct. As noted earlier, this type of item seemed confusing, and the second semester final showed them definitely to be poor discriminators. Table 3 reveals that item 21, the first in Section C, had a - .273 biserial correlation, and item 22 had only a + .116. The section as a whole had an average of + .185. Though these items seem to work well in the face-to-face administration employed in pre-testing, they are evidently confusing and difficult to understand when administered via TV.

The Pearson product moment correlation coefficient between the first semester final and the second semester final was .756.

TABLE 3.

BISERIAL CORRELATIONS OF AND PROPORTION WHO PASSED EACH ITEM ON THE FINAL PRE-TEST AND THE ACTUAL ADMINISTRATION -- SECOND SEMESTER FINAL

Item	Biserial Correlation Coefficient		Proportion Who Passed Item	
	Pre-Test	Final Form	Pre-Test	Final Form
A 1	.371	.732	.887	.867
2	.593	.432	.794	.747
3	.350	.508	.557	.707
4	.564	.312	.309	.480
5	.719	.580	.798	.647
6	.612	.801	.484	.633
7	.614	.742	.876	.900
B 8	.620	.674	.825	.747
9	.431	.721	.649	.780
10	.679	.685	.742	.620
11	.455	.758	.474	.607
12	.407	.635	.691	.733
13	.778	.652	.902	.820
14	.531	.862	.732	.800
15	.511	.627	.515	.640
16	.586	.435	.557	.600
17	.413	.343	.392	.247
18	.589	.664	.660	.753
19	.669	.382	.495	.387
20	.687	.665	.850	.867
C 21	.337	-.273	.680	.433
22	.308	.116	.670	.740
23	.525	.554	.873	.967
24	.349	.305	.567	.633
25	.475	.222	.802	.773
D 26	.449	.384	.639	.740
27	.627	.758	.873	.840
28	.714	.822	.841	.780
29	.658	.479	.809	.787
30	.573	.517	.755	.620
31	.305	.280	.784	.840
32	.360	.542	.670	.613
E 33	.424	.577	.866	.873
34	.692	.621	.841	.733
35	.450	.374	.732	.827
36	.348	.669	.763	.920
37	.454	.411	.495	.327
38	.615	.678	.742	.813
39	.780	.652	.833	.780
40	.416	.263	.753	.900
41	.689	.640	.567	.660
42	.624	.192	.817	.780
43	.340	.555	.515	.593
44	.661	.825	.802	.927
45	.317	.394	.474	.400
46	.475	.520	.639	.667
F 47	.494	.625	.810	.500
48	.608	.649	.810	.553
49	.532	.145	.810	.467
50	.517	.316	.659	.647

3. CONCLUSIONS

Three listening comprehension tests to be administered via TV to fifth graders in their first year of Spanish instruction were developed. Recommendations of experts in modern language teaching and evaluation were followed in the development, and content analysis of the televised Spanish instructional series was made to assure complete coverage of material presented. Rather extensive pre-testing was employed so that satisfactory validity, reliability, discrimination, range of difficulty, and frequency distribution of scores could be achieved.

Table 4 briefly summarizes the tests in terms of average biserial correlation coefficient between each item and total score (the measure of discrimination), average proportion of pupils who answered each item correctly, and split-half reliability (with Spearman-Brown correction).

TABLE 4.

DISCRIMINATION, DIFFICULTY,
AND RELIABILITY OF THREE TESTS

Test	Average Biserial Correlation Coefficient	Average Proportion Who Passed Each Item	Reliability
First Semester Mid-Term	.439	.630	.834
First Semester Final	.540	.720	.899
Second Semester Final	.520	.694	.881

The figures in Table 4 indicate that each test was satisfactory, though both finals were slightly better in terms of discrimination and reliability than the mid-term. Also, the finals were closer to the desired average grade of 70 per cent.

Generally, two types of items were used. The first had pictures on the pupil's answer sheet, and the pupil responded by matching a picture to a statement spoken by the television instructor. The second type of item required choosing one of two possible alternatives (true or false, 1 or 2) to judge a statement or statements spoken by the instructor. In each test the picture items had substantially higher average biserial correlation coefficients than did the others.

The method employed in building these listening comprehension tests, i.e., careful and detailed analysis of course content, exact specification of course objectives, and extensive pre-testing, seems necessary to develop good measuring instruments for elementary school foreign language instruction. The experience in Denver has shown that, although considerable effort is required, good measuring devices can be developed.

APPENDIX A

ANALYSIS OF CONTENT OF THE FIRST SEMESTER'S TELEVISION LESSONS

CONTENT ANALYSIS OF THE FIRST SEMESTER'S TELEVISION LESSONS

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
NOUNS		
Adiós	6	2
agua	37	2
alcoba	30	4
alumna	43	2
alumno	13	7
americano	23	2
amigas	1	5
amiguitos	32	2
amigos	2	12
Antonio	3	3
árbol	Christmas	1
artículos	24	3
banco	41	2
bandera	25	4
baño	38	3
béisbol	20	2
blusa	28	2
boca	17	7
brazos	17	5
bruja	Halloween	3
buho	Halloween	2
cabello	17	5
cabeza	17	3
café	25	10
calabaza	Halloween	2
calcetín	23	3
cama	37	4
camisa	22	5
candela	4	2
carabelas	Columbus	2
cara	21	4
carne	11	2
casa	30	7
Catalina	6	2
cero	45	2
chili	26	3
chocolate	35	3
cientos	21	2
cinco	4	5
cinto	23	3
clase	1	9
cocina	30	6
color	25	4
comedor	30	6
comida	11	5
conversación	36	4

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
NOUNS		
corbata	22	5
cosas	35	3
Cristóbal Colón	13	2
cuarto & cuartos	30	8
cuarto de baño	30	4
cuarto del niño	30	3
cuatro	4	6
cuchara	35	2
cuchillo	35	2
cuello	18	2
cuerpo	17	4
dedos	17	3
día	20	4
días	1	6
dientes	21	3
diez	20	2
dormitorio	37	2
dos	4	9
dulce	Halloween	3
escuela	44	3
Estados Unidos	Christmas	3
estufa	33	4
exterior	30	3
familia	5	19
fiestas	Christmas	3
gato	Halloween	2
globo	45	2
gorra	28	2
gracias	1	9
gusto	5	4
hija	8	10
hijo	8	11
interior	30	2
José	11	3
joven	2	7
juego	20	4
lápiz	20	2
lección	8	9
lechuga	26	3
libro	1	3
Lupe	8	2
madre	8	2
mamá	8	14
manos	17	8
manzana	27	2
mañana	3	6
María	8	4

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
NOUNS		
máscara	Halloween	2
mesa	33	4
mesita	31	5
mexicano	22	2
momento	4	3
muchachas	39	2
muchachos	39	2
muebles	41	2
nada	11	3
nacimiento	Christmas	2
nariz	17	5
Navidad	Christmas	2
nene	10	7
niña	8	3
niñito	5	2
niño	8	6
niños	20	6
noche	4	4
nueve	20	2
número	5	5
ocho	20	2
ojos	17	6
órdenes	31	3
orejas	18	4
padre	8	3
pantalones	23	6
pan	11	2
papá	8	16
papel	27	2
partes	17	4
patio	32	2
pelo	17	6
Pepe	11	2
permiso	4	2
personas	10	6
piano	41	2
piernas	18	4
pies	18	3
pizarra	44	3
platito	34	2
plato	34	5
por favor	1	12
profesor	1	9
profesora	14	2
puerta	30	2
recámara	30	3
refrigerador	30	3
regalo	Christmas	2
ropa	22	6
saco	24	3
sala	6	10
NOUNS		
seis	20	4
señor	2	11
señora	4	7
señorita	2	11
servilleta	35	4
silla	31	6
sobretodo	24	3
sofá	31	4
sombrero	22	7
sopa	35	2
tarde	3	7
taza	34	5
teléfono	36	4
televisión	13	8
tenedor	35	3
tiza	44	3
toalla	39	2
tocador	38	3
todos	3	6
traje	27	4
tres	4	7
uno	4	6
vaso	35	4
ventana	30	5
vestido	24	4
vez	3	5
vista	2	4
vocabulario	45	2
voz	2	2
zapato	23	4
VERBS		
abra	29	3
abran	29	2
abrir	29	2
abro	18	4
aparece	20	2
apunte	2	3
apunten	6	2
bailar	Christmas	2
bebe	10	3
beben	35	3
bebo	35	4
cantan	Christmas	2
cantar	Christmas	2
cerramos	29	2
cerrar	29	2
cierra	29	2
cierran	29	2

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
VERBS		
cierro	18	5
come	17	3
comemos	42	2
comen	Christmas	3
comer	42	2
comiendo	11	5
como	1	18
comprenden	2	2
contar	22	2
decimos	42	2
decir	42	2
deme	20	4
desea	42	3
deseamos	42	3
desean	42	3
desear	42	2
deseo	24	4
dice	4	5
dicen	42	2
dígame	6	4
digo	42	2
dormir	42	2
dormimos	42	2
duerme	20	3
duermen	42	2
duermo	30	3
durmiendo	20	2
entra	29	2
entramos	29	2
entran	29	2
entrar	29	2
entre	30	2
entro	29	2
es	1	44
escribir	20	2
escribo	44	3
escuche	1	3
escuchen	1	3
está	1	39
estamos	16	3
están	2	16
estar	16	2
estás	17	5
estoy	2	7
estudia	42	2
estudiamos	42	2
estudian	18	3
estudio	42	2

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
VERBS		
habla	7	4
hablamos	7	2
hablan	7	2
hablando	20	2
hablar	2	4
hablas	20	2
hablo	7	5
hace	42	2
hacemos	42	2
hacen	42	2
hacer	20	3
haciendo	11	3
hago	42	2
hay	9	11
ir	16	2
juega	29	2
juegan	20	3
jugamos	29	2
jugar	21	3
lavo	38	3
levanta	16	2
levanta[illegible]s	16	2
levantan	16	2
levantar	16	2
levanten	43	2
levántense	43	2
levántese	2	4
levanto	16	4
llama	2	11
mira	13	5
miramos	16	2
miran	16	2
mirando	13	6
mirar	16	2
miro	13	6
pasa	29	2
pasamos	29	2
pasan	29	2
pasar	29	4
pase	4	4
pasen	4	5
paso	29	2
peino	18	4
pintar	41	2
pone	23	4
ponemos	29	2
ponen	29	2
poner	29	2
pongan	35	2
pongo	22	6

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
VERBS		
quita	29	2
quitamos	29	2
quitan	29	2
quitar	29	2
quito	22	4
repasar	6	2
repita	1	3
repitan	1	6
rompe	Christmas	2
seco	38	3
sentado	33	2
sentamos	16	2
sentar	16	2
ser	16	2
sienta	16	3
siéntan	16	3
siente	20	2
siéntese	2	5
siento	16	5
somos	16	4
son	8	15
soy	2	9
tenemos	15	4
tener	42	2
tengo	19	21
tiene	18	14
tienen	39	4
toca	29	2
tocamos	29	2
tocan	29	2
tocar	29	2
toco	29	3
trabaja	33	2
usa	23	4
usamos	42	2
usan	42	2
usar	42	2
uso	23	3
va	16	2
vamos	2	13
van	16	2
vas	20	2
ve	16	2
vemos	16	2
ven	16	2
venga	6	5
vengo	16	2
venimos	16	2
venir	16	2
veo	16	3
ver	2	5

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
VERBS		
viene	16	2
vienen	16	3
visitar	32	2
visto	25	2
vive	30	3
viven	32	2
vivo	30	3
voy	16	7
PRONOUNS		
ella	9	4
me	2	11
mi	3	23
mía	12	2
mis	5	6
qué	6	12
quién	3	11
quiénes	10	3
se	3	15
su	4	10
sus	26	4
suya	17	3
tu	20	2
usted	1	20
ustedes	1	10
yo	2	36
ADVERBS		
acá	6	5
ahora	14	2
allí	4	2
aquí	1	17
así	2	3
bien	1	11
cuántas	10	6
dónde	1	6
hoy	45	4
juntos	3	12
más	4	4
menos	7	2
muy	1	12
no	2	9
presente	44	2
sí	20	5
también	11	6
ADJECTIVES		
alegre	2	2
alta	1	3

Lexical Item	Introduced In Lesson	Number of Lessons in Which It was Emphasized
ADJECTIVES		
azul	25	5
blanca	38	2
blanco	25	4
bonito	24	4
bueno	1	9
cien	21	2
corto	35	2
difícil	2	3
este	1	36
estos	17	7
extraños	Halloween	2
fea	Halloween	2
femenino	8	2
grande	19	11
inteligente	7	3
interesante	20	2
larga	18	2
linda	Christmas	2
lindo	Christmas	2
lista	44	3
masculino	8	2
mucho	5	5
muchas	40	2
negro	25	6
nuevo	45	2
otra	3	5
pequeña	35	2
pequeño	8	4
rojo	25	5
tal	10	2
un	4	11
una	4	15
verdad	2	5
verde	26	3
PREPOSITIONS		
al	20	2
con	1	14
de	5	31
del	9	10
en	3	25
para	24	3
por	3	9
CONJUNCTIONS		
o	11	4
y	1	22
INTERJECTIONS		
hola	6	5
oh	2	3

APPENDIX B

THE FIRST SEMESTER MID-TERM TEST

HABLO ESPAÑOL ©
LEVEL I SPANISH TEST
TEST NUMBER ONE

Part I

In Section A I will make three statements about each picture. One statement, and only one, will be the correct statement for that picture. You will mark your answer by drawing a circle around the "A", "B", or "C", whichever you believe to be the correct answer.

Look at the picture for Example 0 and listen carefully to all three statements.

A. El hijo se pone la camisa.
B. La hija se pone la blusa.
C. Ninguno -- ninguno means neither.

I'll read them once more.

A. El hijo se pone la camisa.
B. La hija se pone la blusa.
C. Ninguno -- (remember that ninguno means neither).

Now circle the "A", the "B", or the "C", whichever you think best describes the picture.....That's right, "B" is correct (La hija se pone la blusa) for Example 0. Remember, if "A" or "B" does not describe the picture, you will circle the "C" to show that neither statement "A" or "B" was correct.

Listen carefully! I will say each statement twice.

Number 1	(A) Es el dedo.	(B) Son los dedos.	(C) Ninguno
Number 2	(A) Son las manos.	(B) Es la mano.	(C) Ninguno
Number 3	(A) Es un niño.	(B) Es una niña.	(C) Ninguno
Number 4	(A) La madre usa un sombrero blanco.	(B) La madre usa un sombrero negro.	(C) Ninguno
Number 5	(A) Es un cinto.	(B) Es un vestido.	(C) Ninguno
Number 6	(A) Pepe mira la televisión.	(B) Pepe come.	(C) Ninguno
Number 7	(A) Es la cabeza.	(B) Es la cara.	(C) Ninguno
Number 8	(A) La clase levanta las manos.	(B) La clase estudia.	(C) Ninguno
Number 9	(A) El pie es grande.	(B) La pierna es grande.	(C) Ninguno
Number 10	(A) Es una señorita.	(B) Es una señora.	(C) Ninguno
Number 11	(A) Es una falda.	(B) Es una corbata.	(C) Ninguno
Number 12	(A) El nene abre los ojos.	(B) El nene juega.	(C) Ninguno
Number 13	(A) La camisa es blanca.	(B) La blusa es negra.	(C) Ninguno
Number 14	(A) Es una camisa.	(B) Es una corbata.	(C) Ninguno
Number 15	(A) Tiene una mano.	(B) Tiene dos manos.	(C) Ninguno

Number 16 (A) La señorita habla. (B) La señorita se viste. (C) Ninguno

Now let's look at Section B.

In this section I will say each statement twice. If the statement is correct, (good Spanish) circle the "sí." If it is incorrect, (poor Spanish) circle "no." Let's do Example 0 of Section B together. "Esta es la madre." Sí o no? Sí, the statement was correct. If I had said "Esta es el madre," it would have been wrong, wouldn't it? So for Example 0 you should have circled "sí" because the statement was correct (or good Spanish). Now let's do Section B. Listen carefully, I will say each statement twice!

Number 1. Esta es la profesora.
Number 2. Esta es el señora.
Number 3. Esta es el niña.
Number 4. Esta es la padre.
Number 5. Esta es la cara.
Number 6. Este es el amigo.
Number 7. Este es el blusa.
Number 8. Estos son los calcetines.

PART II

In Section C I will read a sentence and you must decide which of the three pictures the sentence describes. Remember, the sentence must describe what you can actually see in the picture.

Let's do Example 0 together. Look at the pictures next to Example 0 at the top of the page. "A" shows a little girl wearing a dress. "B" shows a woman wearing a dress. "C" shows a little girl wearing a skirt and blouse. Now listen to the sentence:

"La niña usa una blusa y una falda."

Which picture does the sentence describe?

Sí, picture "C" is correct -- the picture of a girl wearing a blouse and skirt. Each of you should have circled the "C". If you didn't have it circled, circle the "C" in Example 0 now.

Listen carefully! I will read each statement twice.

Number 1. La niña tiene una blusa en la mano.
Number 2. La camisa es blanca.

Number 3. Juan abre la ventana.
Number 4. El joven y sus amigos están en la clase.
Number 5. El niño tiene dos orejas.
Number 6. El joven levanta los brazos.
Number 7. La familia está comiendo.
Number 8. El nene cierra los ojos.
Number 9. El niño tiene dos piernas.
Number 10. La niña cierra la puerta.
Number 11. Hablamos con la boca.
Number 12. María usa una blusa.

Now turn to page 3 and find Section D.

In this section I will say a sentence that requires a response, and then I will give a response. You must then decide if the statement or answer I gave in response is correct.

Now here's Example O of Section D to give you the idea.

"¿Cómo está usted?"......."Así, así."

Is "Así, así" a correct answer for "¿Cómo está usted?" Yes, so you would circle the sí in Example O.

Now let's complete Section D. Listen carefully.......I will say each statement twice.

Number 1.	Cierre la puerta.	Abro la ventana.
Number 2.	¿Cómo está la familia?	Muy bien, gracias.
Number 3.	Ud. habla español muy bien.	Gracias, Juan.
Number 4.	Levántese Ud.	Me levanto.
Number 5.	¿Cómo se llama usted?	Bien gracias y usted?

Now find Section E in the upper right hand quarter of the page.

In Section E I will show you a picture and will say a sentence in Spanish. If what I say tells what is in the picture, circle "sí." If it does not tell what is in the picture, circle "no."

Let's do Example O together (Hold up picture of clothing). "Estos son artículos de ropa." sí or no? Yes, the picture shows "artículos de ropa" so you should circle the sí in Example O. I will say each sentence twice; Look at the picture carefully.

Number 1. Esta es la profesora. (Picture of male professor)
Number 2. Este es un cinto. (Picture of tie)
Number 3. Esta es una blusa. (Picture of a shirt)
Number 4. Esta es la cabeza. (Picture of an arm)

Now let's look at Section F in the lower left hand part of the page. In this section I will make two statements, one statement will be right and one will be wrong. If the first sounds right to you, circle the number one. If the second one sounds right, circle the number two.

Here is Example O:

Statement 1 "La niña como." Statement 2 "La niña come."

Which of these sounds correct? Right, the second one, "La niña come," so you should circle the 2 in Example O.

In this section I will not repeat---I will say each statement once only.

Statement 1 La niña se llama Maria. Statement 2 La niña me llamo María.
Statement 1 Pepe se visten. Statement 2 Pepe se viste.

Now let's look at Section G, the last part of the test. In section G I will give you a sentence and you are to determine if the sentence is correct or if it has a word with a wrong ending. If it is correct as given, mark "sí", but if the sentence is not given correctly, mark "no."

Let's do Example O in section G. "La casa es pardo." sí or no? No, pardo should have been parda, so since an ending was wrong you mark "No" to show the mistake. Listen carefully, I will say each sentence twice.

Number 1. La camisa es blanco.
Number 2. El vestido es azul.
Number 3. La falda es rojo.

Copyright 1960 By
The Denver Public Schools

LEVEL I SPANISH TEST ©

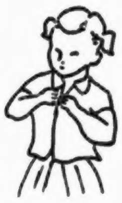
Part 1 - December 7, 1960

Name______________________________ School______________________________

Spanish Teacher______________________________

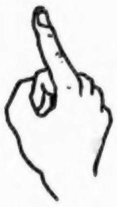
Section A

Example 0.



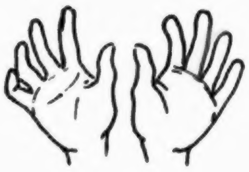
A B C

1.



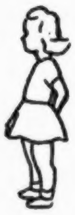
A B C

2.



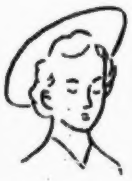
A B C

3.



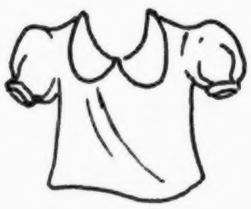
A B C

4.



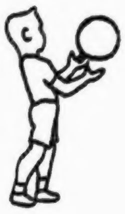
A B C

5.



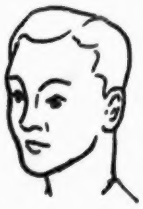
A B C

6.



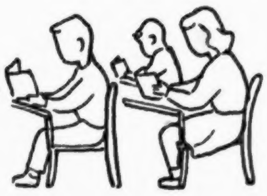
A B C

7.



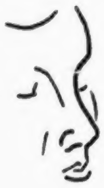
A B C

8.



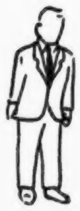
A B C

9.

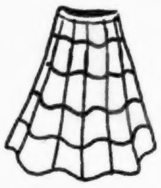


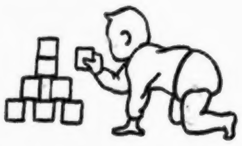
A B C

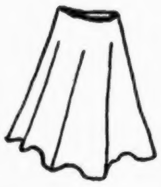
10.

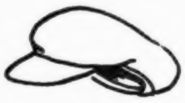


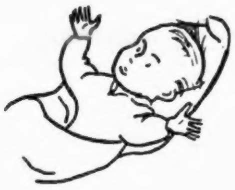
A B C

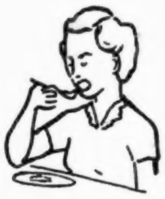
11.  A B C

12.  A B C

13.  A B C

14.  A B C

15.  A B C

16.  A B C

Section B

Example 0. Sí No

1. Sí No

2. Sí No

3. Sí No

4. Sí No

5. Sí No

6. Sí No

7. Sí No

8. Sí No

Copyright 1960 By
The Denver Public Schools

LEVEL I SPANISH TEST ©

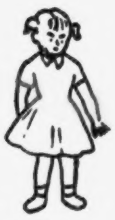
PART 2 - December 9, 1960

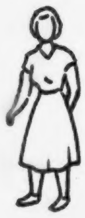
Name________________________________ School__________________________

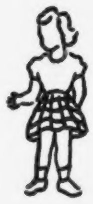
Spanish Teacher_______________________________

Section C

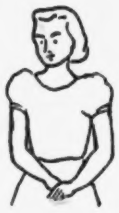
Example 0.

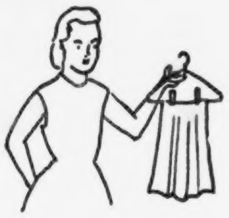

A

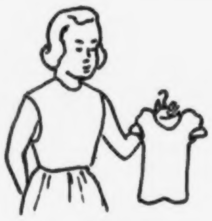

B


C

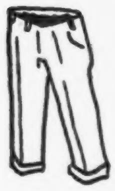
1.

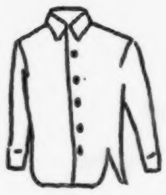

A

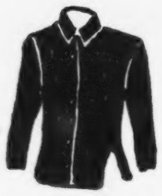

B


C

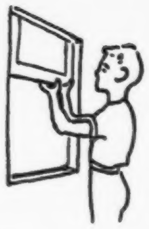
2.

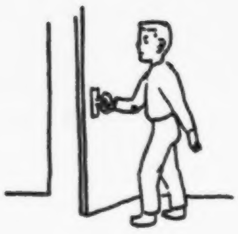

A

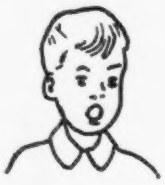

B


C

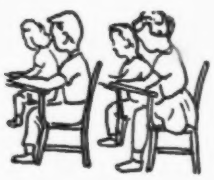
3.

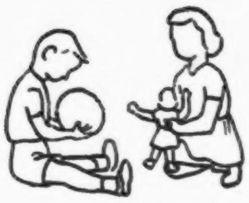

A

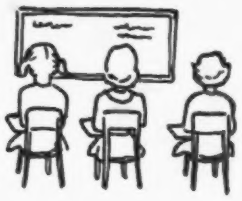

B


C

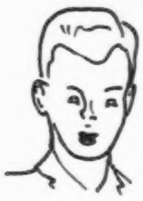
4.

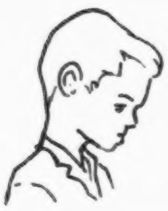

A

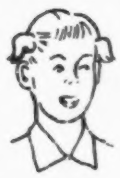

B


C

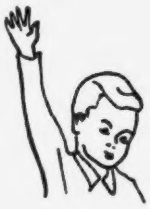
5.

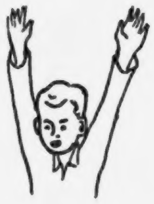

A

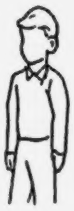

B


C

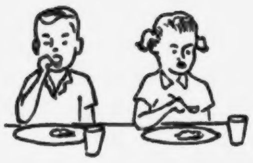
6.

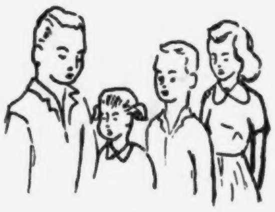

A

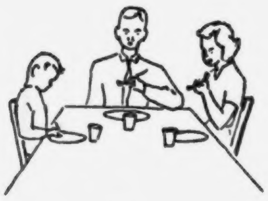

B


C

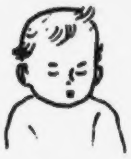
7.

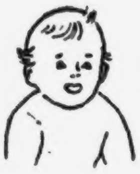

A

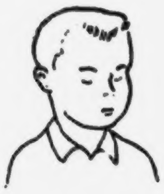

B


C

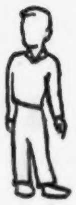
8.

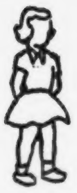

A

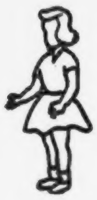

B


C

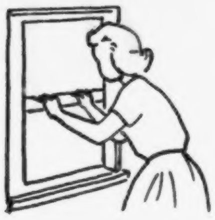
9.

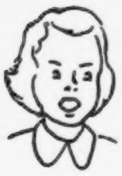

A

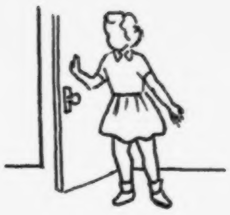

B


C

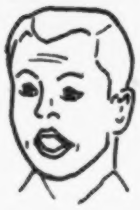
10.

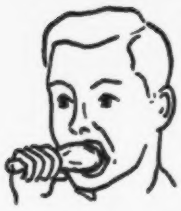

A

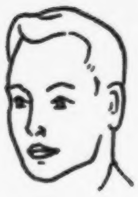

B


C

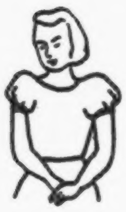
11.

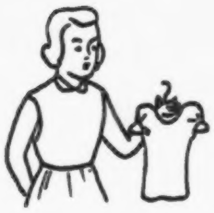

A

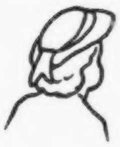

B


C

12.


A


B


C

Section D

Example	0.	Sí	No
	1.	Sí	No
	2.	Sí	No
	3.	Sí	No
	4.	Sí	No
	5.	Sí	No

Section E

Example	0.	Sí	No
	1.	Sí	No
	2.	Sí	No
	3.	Sí	No
	4.	Sí	No

Section F

Example	0.	1	2
	1.	1	2
	2.	1	2

Section G

Example	0.	Sí	No
	1.	Sí	No
	2.	Sí	No
	3.	Sí	No

Copyright 1960 By
The Denver Public Schools

APPENDIX C

THE FIRST SEMESTER FINAL TEST

HABLO ESPAÑOL
LEVEL I SPANISH TEST
TEST NUMBER TWO

Part I

In Section A I will make three statements about each picture. One statement, and only one, will be the correct statement for that picture. You will mark your answer by drawing a circle around the "A", "B", or "C", whichever you believe to be the correct answer. Let's work Example O (at the top of the page) together. Look at the picture for Example O and listen carefully to all three statements.

A. El padre come.
B. La madre come.
C. Ninguno -- ninguno means neither

I'll read them once more.

A. El padre come.
B. La madre come.
C. Ninguno -- (remember that ninguno means neither)

Now circle the "A", the "B", or the "C", whichever you think best describes the picture......that's right, "A" is correct (El padre come) for Example O. Remember, if "A" or "B" does not describe the picture, you will circle the "C" to show that neither statement "A" or "B" was correct.
Listen carefully, I will say each statement twice.

Section A.

1. a. La mano tiene cinco dedos.
   b. La mano tiene dos dedos.
   c. Ninguno.
2. a. Es la nariz.
   b. Es la boca.
   c. Ninguno.
3. a. La madre usa un sombrero blanco.
   b. La madre usa un sombrero negro.
   c. Ninguno.
4. a. Es una camisa.
   b. Es una corbata.
   c. Ninguno.

5. a. Es un niño.
   b. Es una niña.
   c. Ninguno.
6. a. Es una falda.
   b. Es una corbata.
   c. Ninguno.
7. a. Es un cinto.
   b. Es un vestido.
   c. Ninguno.
8. a. El pie es grande.
   b. La pierna es grande.
   c. Ninguno.
9. a. La camisa es blanca.
   b. La blusa es negra.
   c. Ninguno.
10. a. La señorita habla.
    b. La señorita se viste.
    c. Ninguno.
11. a. Es una señorita.
    b. Es una señora.
    c. Ninguno.
12. a. Es el dedo.
    b. Son los dedos.
    c. Ninguno.

Now look at Section B. In Section B I will read a sentence and you must decide which of the three pictures the sentence describes. Remember, the sentence must describe what you can actually see in the picture.

Let's do Example 0 together. Look at the picture next to Example 0 near the top of page 2. "A" shows a family of five. "B" shows five boys talking together. "C" shows five girls talking together. Now listen to this sentence: "En esta familia hay cinco personas." Which picture does the sentence describe? Yes, picture "A" is correct -- the picture of a family of five.

Listen carefully, I will read each statement twice!

Section B.

13. La niña usa una blusa y una falda.
14. Las zapatos son negros.
15. El padre se pone zapatos.
16. El hijo estudia.
17. La familia está comiendo.
18. La camisa es blanca.
19. El joven levanta los brazos.
20. La niña tiene una blusa en la mano.
21. El niño tiene dos orejas.
22. Habla con la boca.
23. El niño tiene dos piernas.
24. Juan abre la ventana.
25. María usa una blusa.

PART II

In Section C I will read a sentence and you must decide which of the four pictures the sentence describes. Remember, the sentence must describe what you can actually see in the picture.

Let's do Example O of Section C together. Look at the picture next to Example O at the top of the page. "A" shows a boy using a knife. "B" shows a boy using a fork. "C" shows a boy using a spoon, and "D" shows him wiping his mouth with a napkin.

Now listen to the sentence: "El niño usa una cuchara."

Which picture does the sentence describe?

Yes, picture "C" is correct -- the picture of a boy using a spoon.

Listen carefully, I will read each statement twice!

Section C.

26. Hay una mesa y seis sillas en este cuarto.
27. La familia está en la sala.
28. El niño habla por teléfono con su amigo.
29. El niño usa un vaso para beber.
30. En este cuarto duermo.
31. En este cuarto come la familia.
32. La madre se lava las manos en la cocina.
33. La madre cierra la ventana.

34. La niña se lava la cara.
35. El nene tiene un cuchillo en la mano.

Now look at Section D. I will say each statement twice. If the statement is correct (good Spanish), circle the "si." If it is incorrect (poor Spanish), circle the "no."

Let's do Example 0 of Section D together. "Esta es el cabeza." Sí o no? No, the statement is not correct. If I had said "Esta es la cabeza" it would have been correct, wouldn't it?

Now let's do Section D together. Listen carefully, I will say each statement twice.

Section D.

36. Este es el hijo. sí no
37. Esta es la silla. sí no
38. Esta es la servilleta. sí no
39. Esta es el cuchara. sí no
40. Este es el vaso. sí no
41. Este es la comedor. sí no
42. Hablamos con el boca. sí no
43. Esta es la cara. sí no
44. Este es el blusa. sí no

Now look at Section E -- the last section of the test.

In this section I will show you a picture and will say a sentence in Spanish. If what I say tells what is in the picture, circle "sí." If it does not tell what is in the picture, circle "no."

Let's do Example 0 together. (Hold up a picture of a spoon.)

"Esta es un tenedor." Sí o no? No, the picture shows una cuchara, no un tenedor, so you should circle "no" for Example 0. I will say each sentence twice. Look at each picture carefully.

Section E.

45. Esta es la cabeza. (Picture of arm) sí no
46. Esta es una blusa. (Picture of skirt) sí no
47. Este cuarto es la sala. (Picture of living room) sí no
48. Esta es una puerta. (Picture of door) sí no
49. Esta es una mesita. (Picture of a night table) sí no
50. Esta es la profesora. (Picture of male professor) sí no

Copyright 1961 By
The Denver Public Schools

LEVEL I SPANISH TEST ©

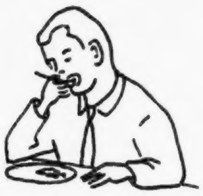
Part 1 - January 23, 1961

Name________________________________ School________________________

Spanish Teacher________________________________

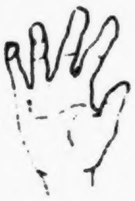
Section A

Example 0.



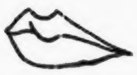
A B C

1



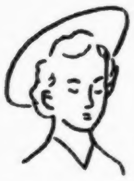
A B C

2



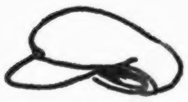
A B C

3



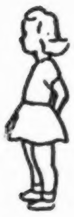
A B C

4



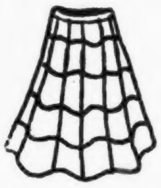
A B C

5



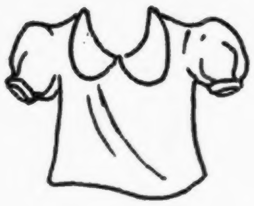
A B C

6



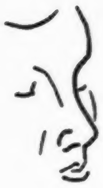
A B C

7



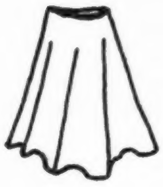
A B C

8



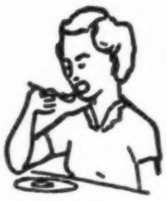
A B C

9

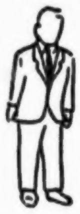


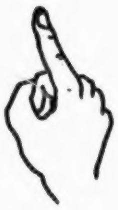
A B C

10



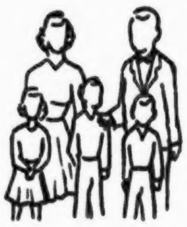
A B C

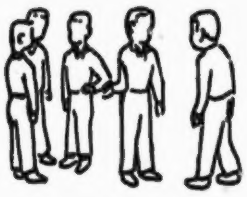
11  A B C

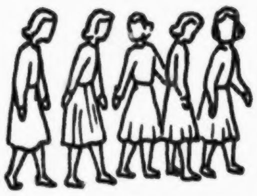
12  A B C

Section B

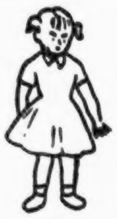
Example 0.

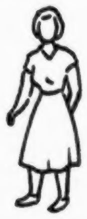

A

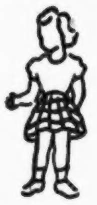

B


C

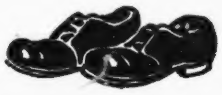
13

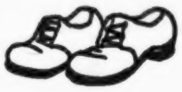

A

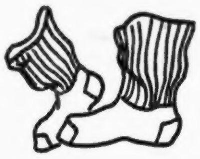

B


C

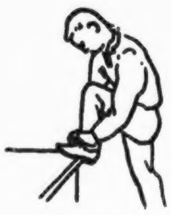
14

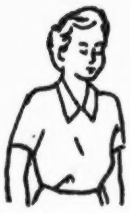

A

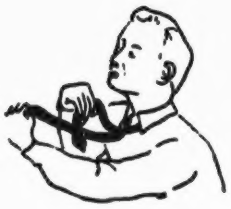

B


C

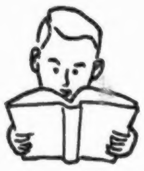
15

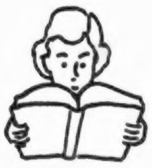

A

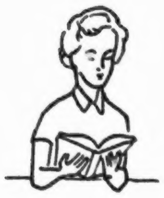

B


C

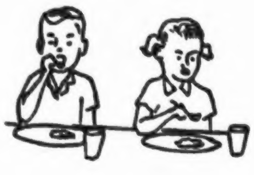
16

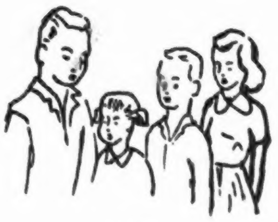

A

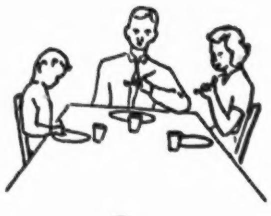

B


C

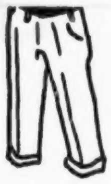
17

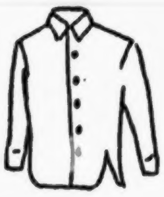

A

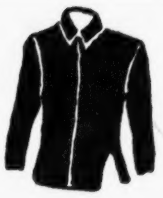

B


C

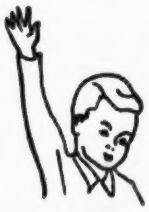
18

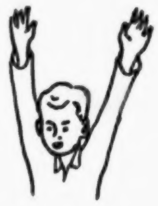

A

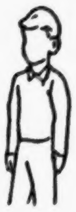

B


C

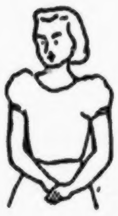
19

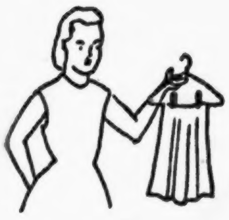

A

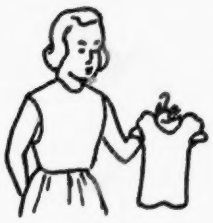

B


C

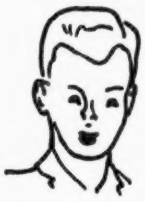
20

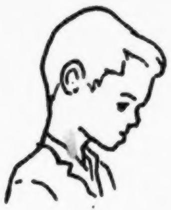

A

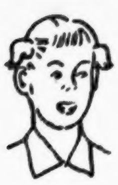

B


C

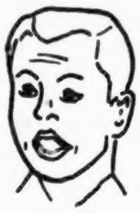
21

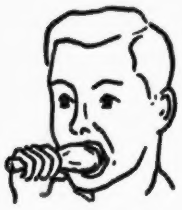

A

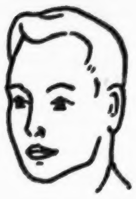

B


C

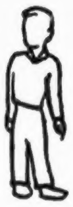
22

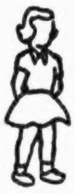

A

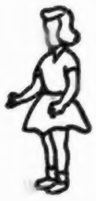

B


C

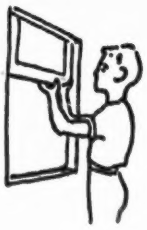
23

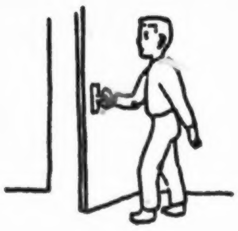

A

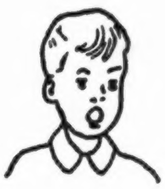

B


C

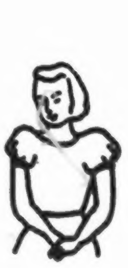
24

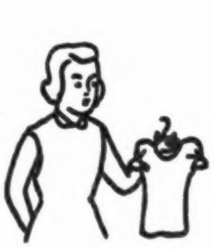

A

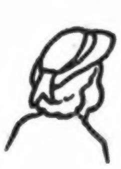

B


C

25


A


B


C

Copyright 1962
By The
Denver Public Schools

LEVEL I SPANISH TEST©

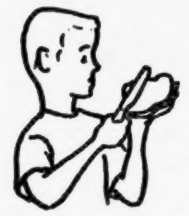
PART 2 - January 25, 1961

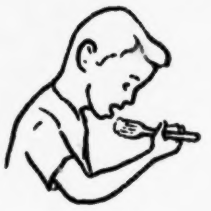
Name________________________________ School________________________

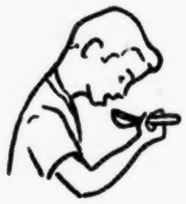
Spanish Teacher_______________________________

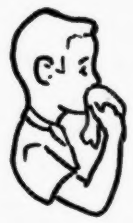
Section C

Example 0.

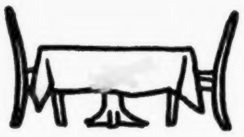

A

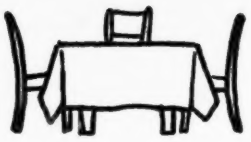

B

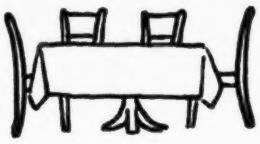

C

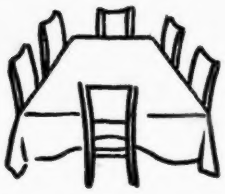

D

26

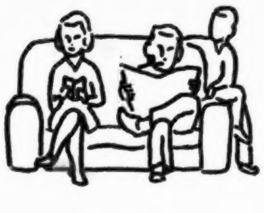

A

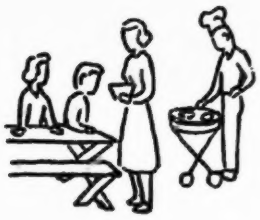

B

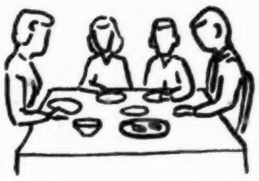

C

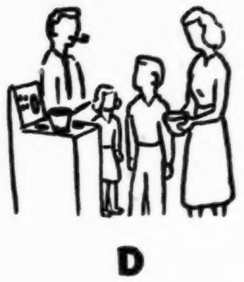

D

27

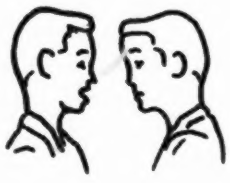

A

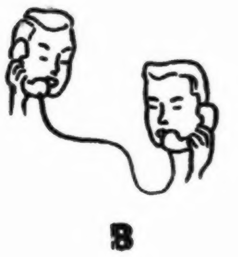

B

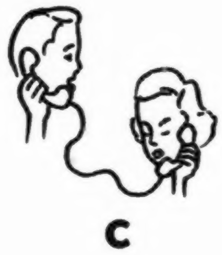

C

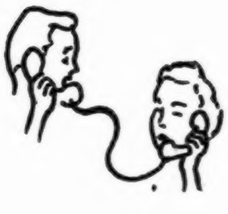

D

28

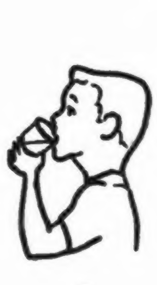

A

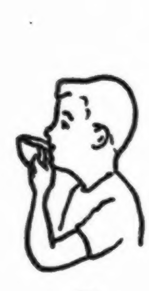

B

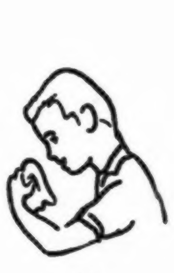

C

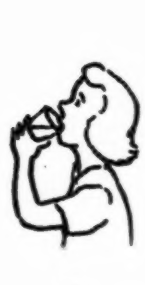

D

29

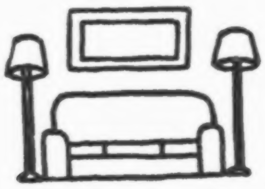

A


B

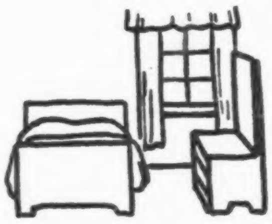

C


D

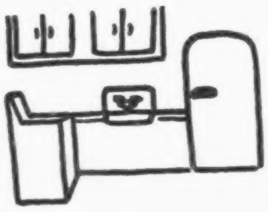
30



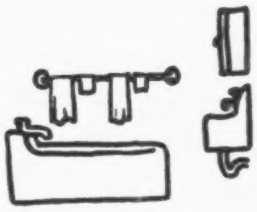
A



B

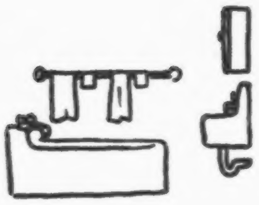


C

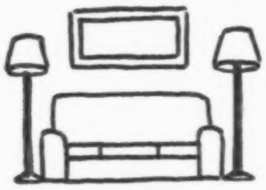


D

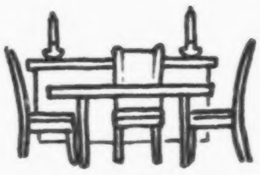
31



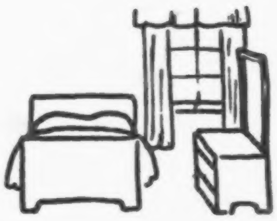
A



B

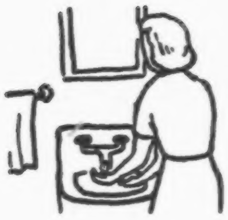


C

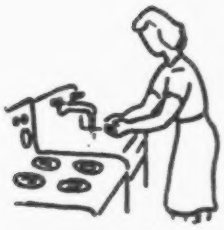


D

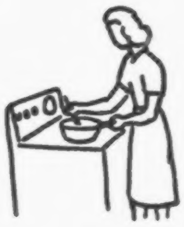
32



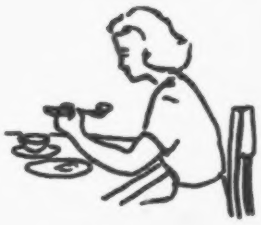
A



B

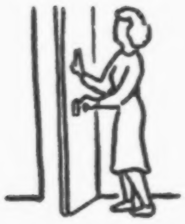


C

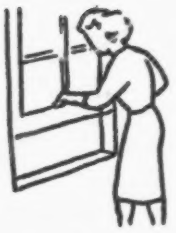


D

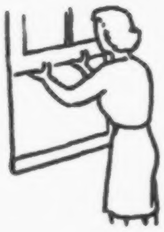
33



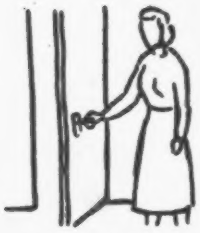
A



B

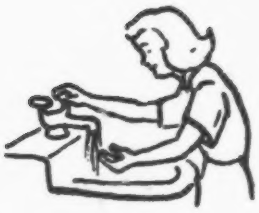


C

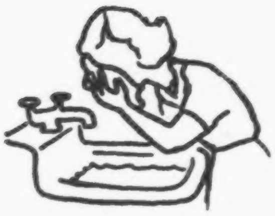


D

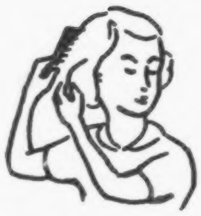
34



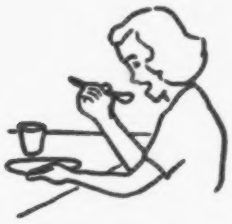
A



B

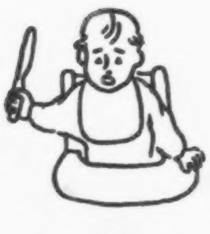


C

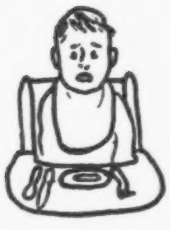


D

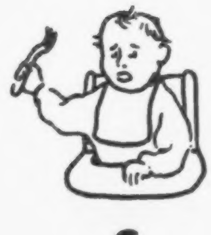
35



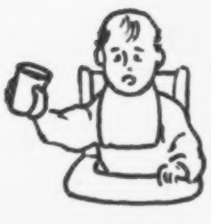
A



B



C



D

Section D

Example 0. Sí No

36. Sí No

37. Sí No

38. Sí No

39. Sí No

40. Sí No

41. Sí No

42. Sí No

43. Sí No

44. Sí No

Section E

Example 0. Sí No

45. Sí No

46. Sí No

47. Sí No

48. Sí No

49. Sí No

50. Sí No

Copyright 1961 By
The Denver Public Schools

APPENDIX D

THE SECOND SEMESTER FINAL TEST

HABLO ESPAÑOL ©
LEVEL I SPANISH TEST
TEST NUMBER THREE

Part I

In Section A I will read a sentence and you must decide which of the three pictures the sentence describes. Remember, the sentence must describe what you can actually see in the picture.

Look at the pictures next to Example 0. "A" shows 12:15. "B" shows 12:45. "C" shows 2:15. Now listen to this sentence: "Son las doce y quince." Which picture does the sentence describe? Yes, picture "A" is correct -- the picture showing 12:15. Each of you should have circled the "A." If you don't have "A" circled, circle the "A" in Example 0 now.

Listen carefully! I will read each statement twice!

Section A.

1. La niña va a la escuela.
2. Todos los alumnos escriben.
3. Los hijos toman leche para el desayuno.
4. Mamá compra comida para el desayuno.
5. La mamá prepara la comida.
6. La mamá tiene una botella.
7. Los niños van a la escuela.

In Section B I will read a sentence and you must decide which of the four pictures the sentence describes. Remember, the sentence must describe what you can actually see in the picture.

Look at the pictures next to Example 0. "A" shows a boy and a girl playing with a ball. "B" shows a boy playing with a ball and a girl jumping rope. "C" shows a boy and a girl reading, and "D" shows a baby playing with blocks. Now listen to the sentence: "Los alumnos juegan con una pelota." Which picture does the sentence describe? Yes, picture "A" is correct -- the picture of a boy and a girl playing with a ball. Each of you should have circled the "A." If you don't have it circled, circle the "A" in Example 0 now.

Listen carefully! I will read each statement twice!

Section B.

8. El niño escribe en el papel con un lápiz.
9. Papá compra una camisa pequeña para su hijo.
10. La mamá sirve carne.
11. La mamá y los niños preparan el almuerzo.
12. La familia está comiendo en la cocina.
13. La niña mira a un calendario.
14. El vendedor vende ropa.
15. Compra una botella de leche.
16. La familia está cenando en el comedor.
17. Ella tiene una falda corta.
18. Son las tres y media.
19. Es la primavera y hace buen tiempo.
20. El papá compra una corbata y una camisa.

Now look at Section C. In this section I am going to read two statements - one statement is correct, the other one is incorrect. Remember, one statement will be right, one will be wrong. If the first one sounds right to you circle number one. If the second one sounds right circle number two.
Listen to the two sentences for Example 0. Number 1: "La alumna lee." Number 2: "La alumna leen." Which statement is correct? Yes, number 1. "La alumna lee." Circle number 1.
Listen carefully for I will say each statement only once.

Section C.

21. 1 Las naranjas está en la tienda.
    2 Las naranjas están en la tienda.
22. 1 La ropa es barata.
    2 La ropa es barato.
23. 1 El alumno escribe en la pizarra.
    2 El alumno escribimos en la pizarra.
24. 1 El padre usa el azúcar.
    2 El padre usamos el azúcar.
25. 1 Muchas personas compra ropa.
    2 Muchas personas compran ropa.

Part II

In Section D I will read a sentence and you must decide which of the three pictures the sentence describes. Remember, the sentence must describe what you can actually see in the picture.

Look at the picture next to Example O. "A" shows a boy playing with a ball and a girl jumping rope. "B" shows a girl looking at a boy holding a book on his head. "C" shows boys playing basketball. Now listen to this sentence: "Los niños juegan juegos." Which picture does the sentence describe? Yes, picture "A" is correct -- the picture showing the children playing two separate games. Each of you should have circled the "A." If you don't have "A" circled, circle the "A" in Example O now.

Listen carefully! I will read each statement twice!

Section D.

26. Los niños cantan canciones.
27. La escuela tiene muchas ventanas.
28. La mamá compra ropa.
29. Ponen platos solamente.
30. El niño está contento.
31. Los alumnos salen de la escuela.
32. Las niñas dan regalos.

In Section E I will make three statements about each picture. One statement, and only one, will be the correct statement for that picture. You will mark your answer by drawing a circle around the "A", "B", or "C", whichever you believe to be the correct answer.

Look at the picture for Example O and listen carefully to all three statements.

A. Es un alumno.
B. Es una alumna.
C. Ninguno (ninguno means neither).

I'll read them once more.

A. Es un alumno.
B. Es una alumna.
C. Ninguno (remember that ninguno means neither).

Now circle the "A", the "B", or the "C", whichever you think best describes the picture ... that's right, "B" is correct ("Es una alumna") for Example O. Remember, if "A" or "B" does not describe the picture, you will circle the "C."

Listen carefully! I will say each statement twice!

Section E.

33. A. Son libros.
B. Es libro.
C. Ninguno.
34. A. Es una ventana.
B. Es una semana.
C. Ninguno.
35. A. El gato es blanco.
B. El plato es blanco.
C. Ninguno.
36. A. El globo está en la mesa.
B. La bandera está en la mesa.
C. Ninguno.
37. A. Es el número veinte y uno.
B. Es un número.
C. Ninguno.
38. A. Son las ventanas.
B. Son las puertas.
C. Ninguno.
39. A. Escribe en el papel.
B. Escribe en la pizarra.
C. Ninguno.
40. A. Es nueve.
B. Es jueves.
C. Ninguno.
41. A. Es jueves.
B. Son huevos.
C. Ninguno.
42. A. Es jugo de tomate.
B. Es jugo de naranja.
C. Ninguno.
43. A. La taza es blanca.
B. La tiza es blanca.
C. Ninguno.
44. A. Es la bandera.
B. Es la primavera.
C. Ninguno.

45. A. Es un veinte.
    B. Hace viento.
    C. Ninguno.

46. A. Hace fresco.
    B. Es un refresco.
    C. Ninguno.

In Section F, I will say each statement twice. If the statement is correct (good Spanish), circle the "Sí." If it is incorrect (poor Spanish), circle the "No."

Let's do Example 0 of Section F together. "Esta es el escuela." Sí o no? No, the statement is not correct. If I had said "Esta es la escuela" it would have been correct.

So, for Example 0 of Section F you should have "No" circled. If you don't have "No" circled, do it now.

Section F.

47. Esta es el sala de clase.	Sí	No
48. Es la estación de vacaciones.	Sí	No
49. Las padres compran ropa.	Sí	No
50. Los huevos son blancos.	Sí	No

Copyright 1961 By
The Denver Public Schools

LEVEL I SPANISH TEST ©

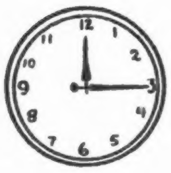
PART I – MAY 22, 1961

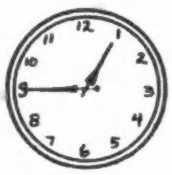
NAME________________________ SCHOOL________________________
SPANISH TEACHER________________________

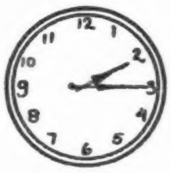
* * * *

SECTION A

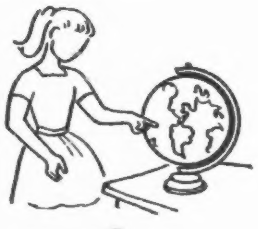
EXAMPLE 0

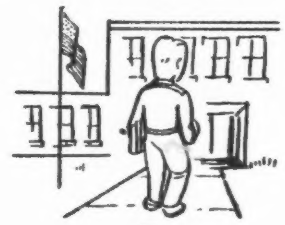

A

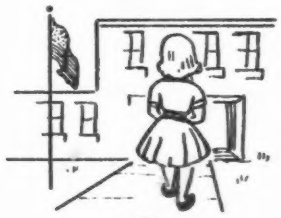

B


C

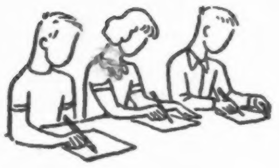
1

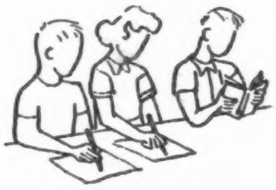

A

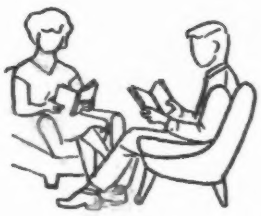

B


C

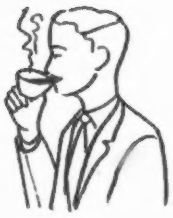
2

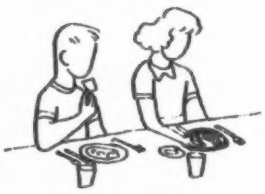

A

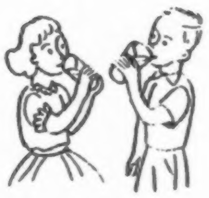

B


C

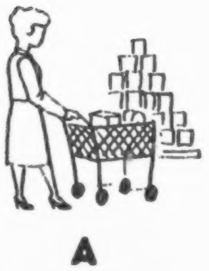
3

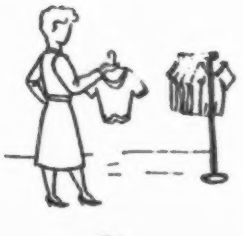

A

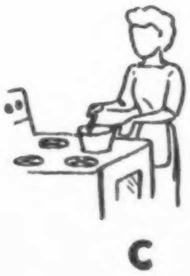

B


C

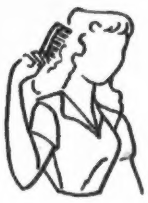
4

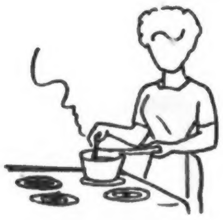

A

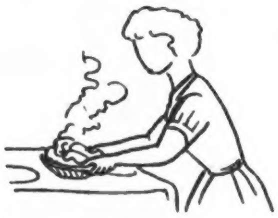

B


C

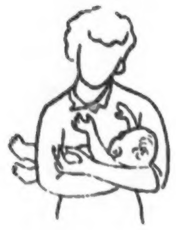
5

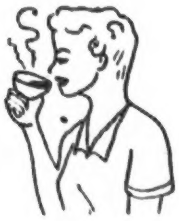

A

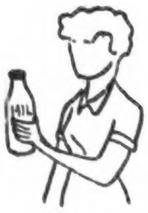

B


C

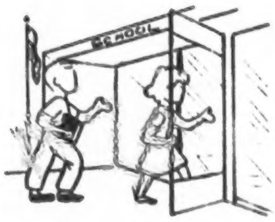
6


A

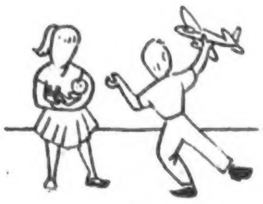

B

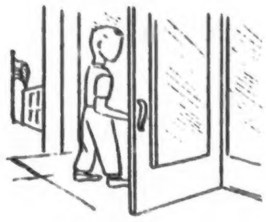

C

7



A

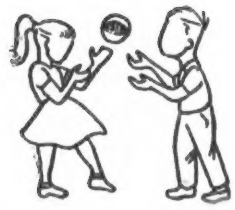

B

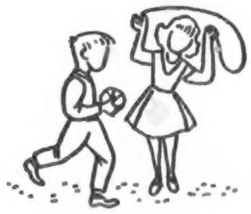

C

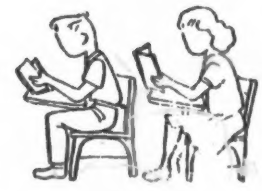
* * * * *

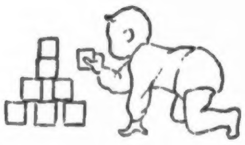
SECTION B

Example 0

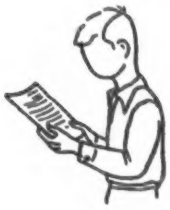

A

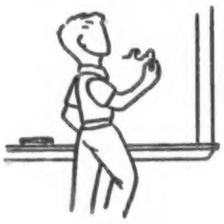

B

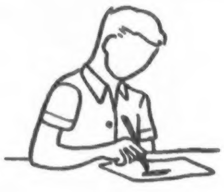

C

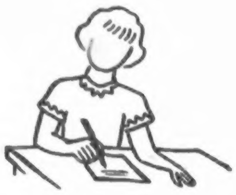

D

8

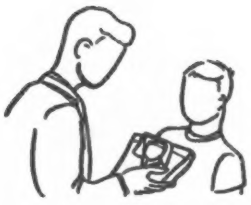

A

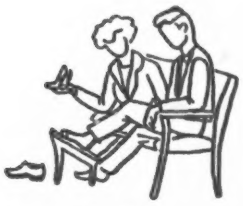

B

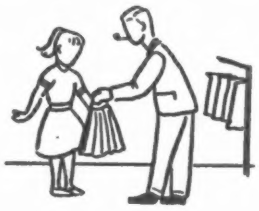

C

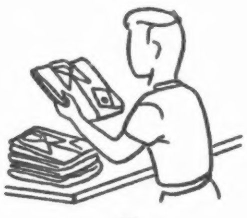

D

9

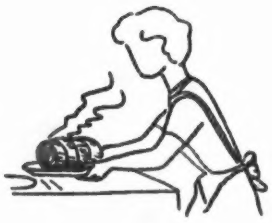

A

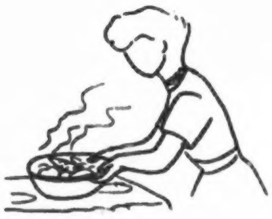

B

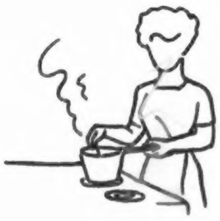

C

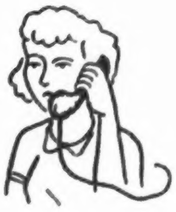

D

10

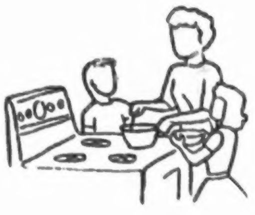

A

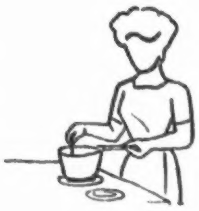

B

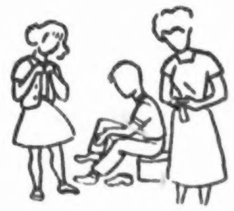

C

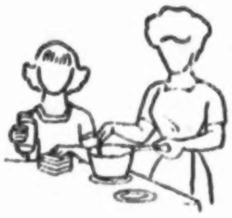

D

11

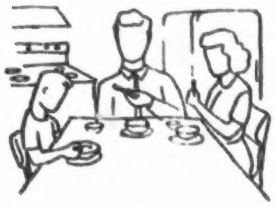

A

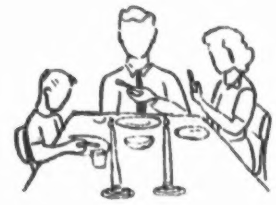

B

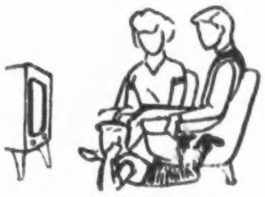

C

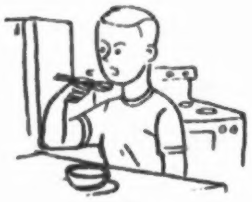

D

12

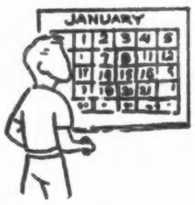

A


B

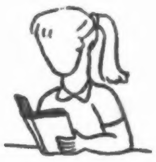

C

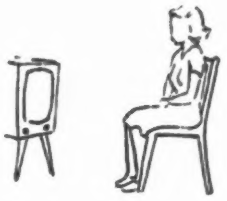

D

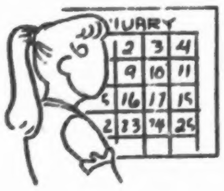
13



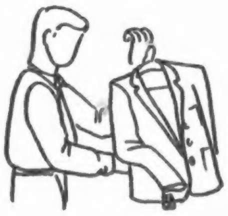
A

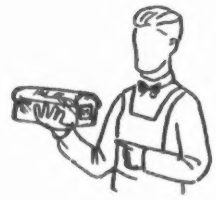

B

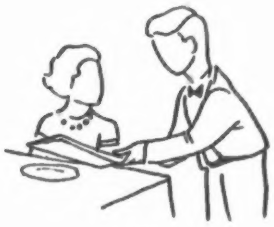

C

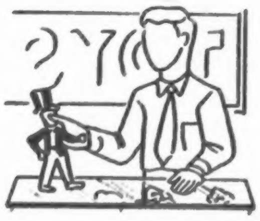

D

14

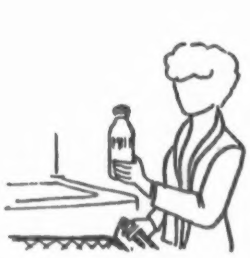

A

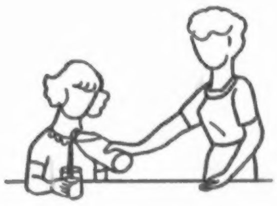

B

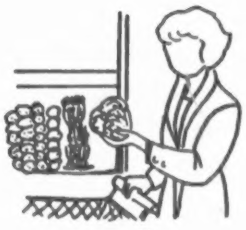

C

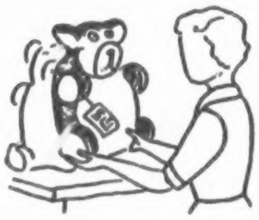

D

15

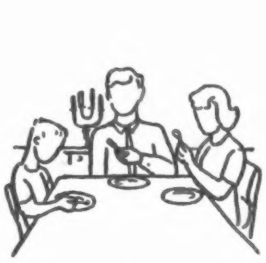

A

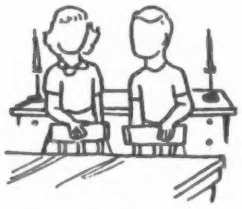

B

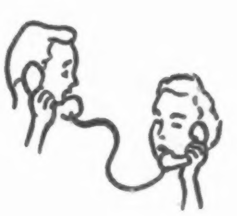

C

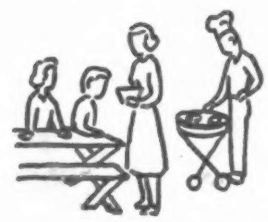

D

16

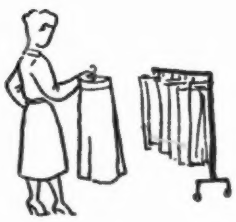

A

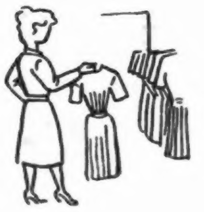

B

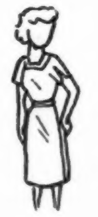

C

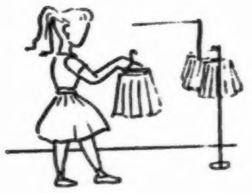

D

17

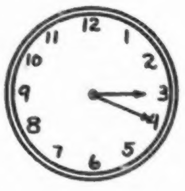

A

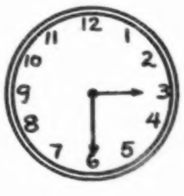

B

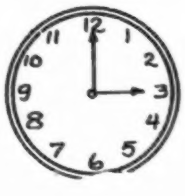

C

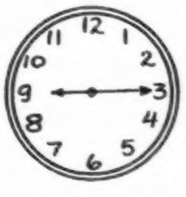

D

18

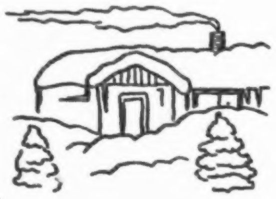

A

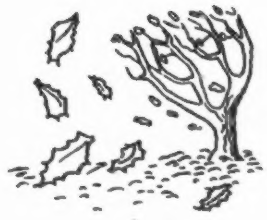

B

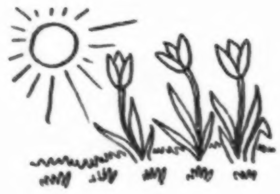

C

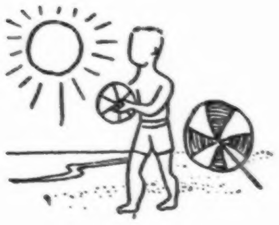

D

19

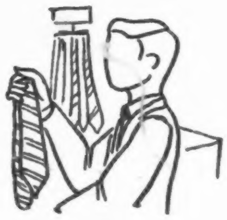

A

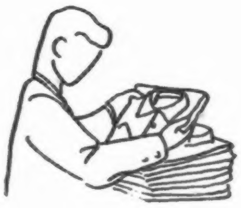

B

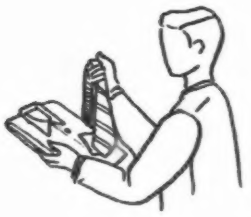

C

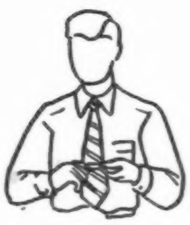

D

20


A


B


C


D

* * * * *

SECTION C

Example 0 1 2

21 1 2

22 1 2

23 1 2

24 1 2

25 1 2

COPYRIGHT 1961 BY THE
DENVER PUBLIC SCHOOLS
DENVER, COLORADO

LEVEL I SPANISH TEST ©

PART 2 - MAY 24, 1961

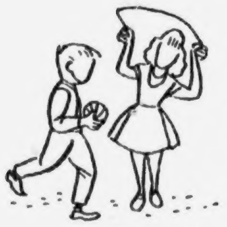
NAME_________________________ SCHOOL_________________________

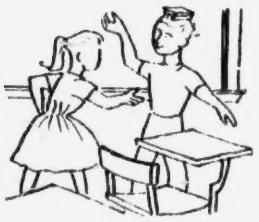
SPANISH TEACHER_________________________

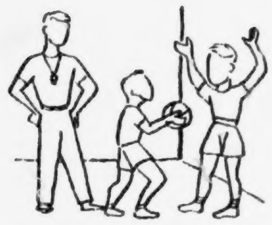
* * * *

SECTION D

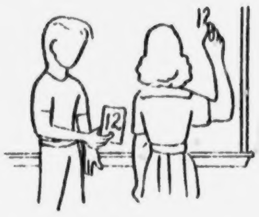
EXAMPLE 0


A

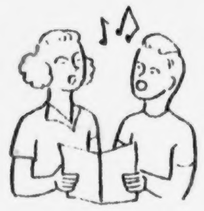

B

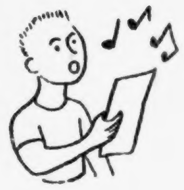

C

26

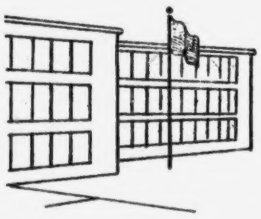


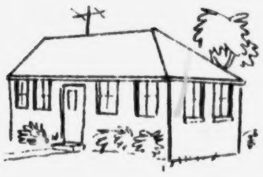
A

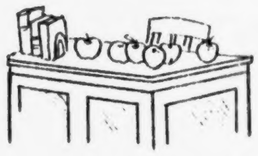

B


C

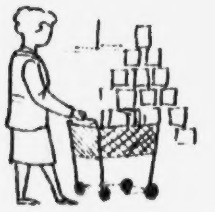
27

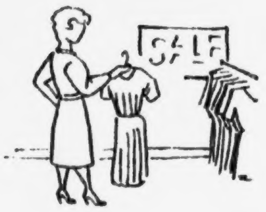

A


B

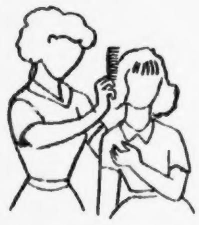

C

28

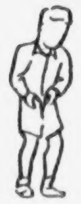

A

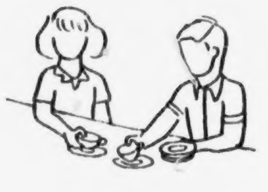


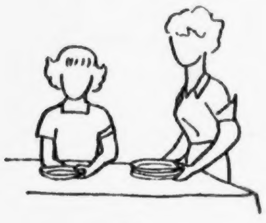
B


C

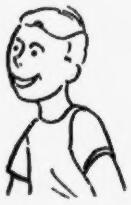
29


A

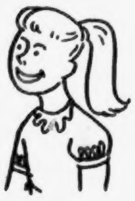

B


C

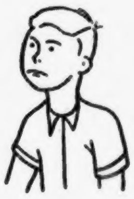
30



A

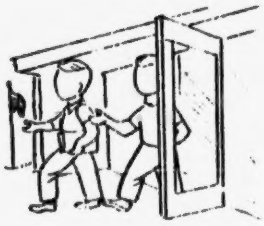


B

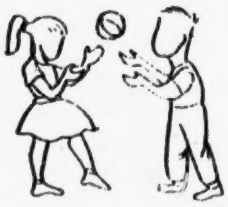


C

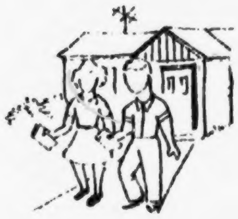
31



A

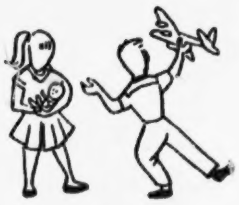


B

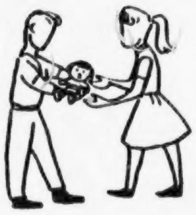


C

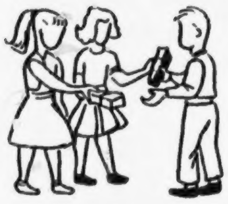
32



A



B

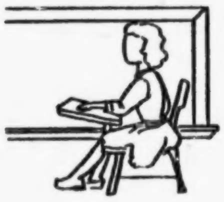


C

* * * * *

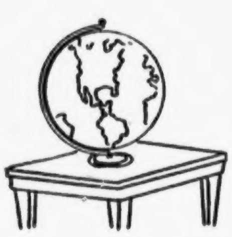
SECTION E

Example O



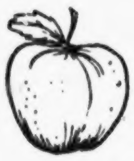
A B C

33



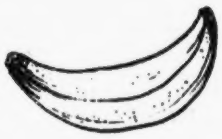
A B C

34



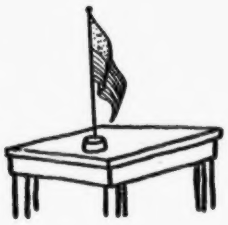
A B C

35



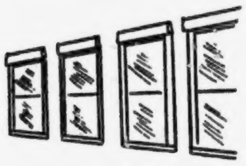
A B C

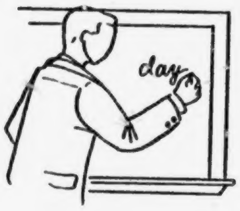
36



A B C

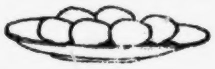
37 20 A B C

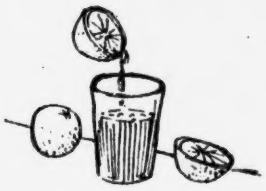
38  A B C

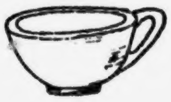
39 

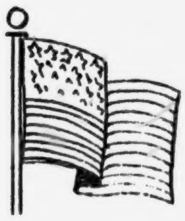
A B C

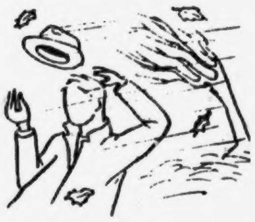
40 9 A B C

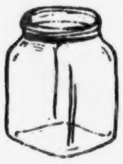
41  A B C

42  A B C

43  A B C

44  A B C

45  A B C

46  A B C

* * * * *

SECTION F

Example O Sí No

47 Sí No

48 Sí No

49 Sí No

50 Sí No

COPYRIGHT 1961 BY THE
DENVER PUBLIC SCHOOLS
DENVER, COLORADO

REFERENCES

Andersson, Theodore, Anges Brady, et. al. *Foreign Languages in Elementary Schools, Some Questions and Answers*. A report of two work conferences sponsored by the Modern Language Association. New York: Modern Language Association, 1954. 21 pp.

Brooks, Nelson. *Language and Language Learning*. New York: Harcourt, Brade, and Company, 1960. 238 pp.

Guilford, J. P. *Psychometric Methods*. New York: McGraw-Hill Book Company, Inc. 597 pp.

Huebener, Theodore. *How to Teach Foreign Languages Effectively*. New York: New York University Press, 1959. 198 pp.

Keesee, Elizabeth. *Modern Foreign Languages in the Elementary School*. Washington, D.C.: U. S. Department of Health, Education, and Welfare, 1960. 65 pp.

MacRae, Margit W. *Teaching Spanish in the Grades*. Boston: Houghton Mifflin Company, 1957. 408 pp.

Peloro, Filomena C., Alexander S. Hughes, J. Donald Bowen, et. al. *Northeast Conference on Teaching of Foreign Languages*. Report of Working Committee III -- Elementary and Junior High School Curricula. 1959. 14 pp.

Selvi, Arthur M., Morton Briggs, Rudolph V. Oblom, et. al. *Foreign Languages Grades 7 -- 12*. Curriculum Bulletin Series, No. V. Hartford, Connecticut: State Department of Education, 1958. 61 pp.

OTHER WORKS HELPFUL IN COURSE DEFINITION AND TEST CONSTRUCTION

Babcock, Edna E. "Foreign Languages in a Changing Curriculum." *Hispania*. November 1952, 35. pp. 431-432.

Ellert, E. E. and L. V. Ellert. "Teaching Modern Languages to Elementary School Children." *Educational Research Bulletin*. January 1953, 32, No. 1. pp. 1-6, 27-28.

Hill, Archibald T. *Introduction to Linguistic Structure*. New York: Harcourt, Brace and Company, 1958.

Lado, Robert. *Linguistics Across Cultures -- Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press, 1957.

ED035269

EM007066

DEVELOPMENT OF PROGRAMED LEARNING MATERIALS FOR USE WITH TELEVISED SPANISH INSTRUCTION

- DELBERT L. BARCUS
- JOHN L. HAYMAN, JR.
- JAMES T. JOHNSON, JR.

Denver Public Schools • Stanford University

RESEARCH ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

The research reported herein was supported by a grant from the United States Office of Education, Department of Health, Education, and Welfare.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED035269

DEVELOPMENT OF PROGRAMED LEARNING MATERIALS FOR USE WITH TELEVISED SPANISH INSTRUCTION

by

Delbert Barcus
John L. Hayman, Jr.
James T. Johnson, Jr.

DENVER-STANFORD PROJECT ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

****

School District Number One
City and County of Denver
Denver, Colorado

Institute for Communication Research
Stanford University
Stanford, California

Report Number 4
November 1961

SUMMARY

The Problem

First year pupils in the Denver-Stanford project study Spanish by the audio-lingual method. During their second year of instruction, however, they learn to read and write the vocabulary and functional grammar used orally the first year.

If programed materials could be developed that would teach reading and writing as well as or better than the conventional teacher-directed approach, the serious shortage of trained teachers and the resulting administrative scheduling difficulties could be overcome.

Results

Both programed and conventional texts were prepared so that a comparison could be made. The lack of prior experience or reported results in the area made programing a time consuming and exacting process. It meant that trials with a wide variety of pupils were necessary.

The literature related to programing was thoroughly examined, and the programer, a language specialist with years of teaching experience, very carefully defined the exact objectives to be achieved. However, this preparation proved highly inadequate. The procedures to be followed could be determined only after many hours of work with pupils.

This report describes the process of development which evolved and gives preliminary results of effectiveness. Large scale trials of the programed materials and comparisons with conventional instruction will be conducted during the 1961-62 school year.

The preliminary results show that pupils learned a significant amount from the programed materials. They also suggest that classroom factors, independent of the materials and of pupil ability, affect the amount learned.

Development of Programed Learning Materials for use with Televised Spanish Instruction

by

Delbert Barcus
John L. Hayman, Jr.
James T. Johnson, Jr.

The Denver Public Schools and Stanford University's Institute for Communication Research are currently engaged in a joint research project on the context of instructional television. The purpose of the project is to learn how instructional television can best fit into the total teaching situation. A substantial amount of research has established that television is a very effective teaching medium. Ways of combining it with other educational activities must now be considered, and the Denver-Stanford project is a beginning effort in this direction. Kenneth E. Oberholtzer is principal investigator for the Denver Public Schools and Wilbur Schramm is principal investigator for Stanford University. This is one of a number of project progress reports.

The Denver-Stanford project is concerned with teaching beginning Spanish to some 12,000 fifth and sixth grade pupils. All pupils view televised Spanish lessons and engage in teacher-directed classroom activities designed to complement the TV lessons. Three half-hour periods per week are devoted to Spanish instruction.

In keeping with current practice in elementary school foreign language instruction, audio-lingual practice is used exclusively during the first year of study (fifth grade). During the second year (sixth grade) audio-lingual skills are taught primarily, but reading and writing are also introduced. The vocabulary and functional grammar used in reading and writing are limited to material which is taught audio-lingually in the fifth grade and with which pupils have oral facility. Both programed and conventional texts were developed during the 1960-61 school year for this aspect of Spanish instruction.

Acquisition of reading and writing skills through programed materials and through conventional teacher-directed methods will be compared during the 1961-62 school year, with some 6,000 sixth graders taking part. Three versions of the programed materials will be used. Two of them involve texts in which different sequences of frames produce different error rates; the third uses the programed materials in a teaching machine. One objective, of course, is to learn as much about the use of automated materials as possible. But, in keeping with the fundamental purpose of the project, the ultimate question to be considered here is, "How can educational television and automated instruction be combined to produce optimum learning?"

This report discusses development of the automated instruction program and presents preliminary test results.

I. Program Development

a. Determination of Programing Procedures

A careful study of available information on automated learning was made prior to any programing. Especially useful were Eugene Galanter's Automatic Teaching: The State of the Art (4); "Teaching Machines: an Annotated Bibliography" by Edward B. Fry, Glenn L. Bryan, and Joseph W. Rigney (3); and Teaching Machines and Programmed Learning by A. A. Lumsdaine and Robert Glaser (5).

The following general steps for developing the program were then established:

1. Outline the behavioral outcomes to be achieved.
2. Analyze the elementary components necessary to achieve such behavior.
3. List the heuristic elements.
4. Prepare a tentative set of frames on 3 x 5 cards.

5. Test the frames on individual sixth graders.
6. Revise the frames.
7. Test the frames on groups of pupils.

These limitations were to be observed:

1. Translations were to be avoided.
2. Visuals were to be used, when possible, to establish the meaning of words.
3. Grammatical constructions were to be taught by induction rather than by rules.
4. The program was to be of the linear, error-free type.

The general behavioral outcomes to be achieved by the end of the year were defined as follows:

1. The pupil should be able to understand the meaning of written Spanish words, sentences, and paragraphs composed of words and phrases used audio-lingually in Level I (fifth grade).
2. The pupil should be able to write any sentence in Spanish composed of words and phrases used audio-lingually in Level I.

Concerning outcome 1, all words are taught in context; consequently the pupils would be expected to recognize the meaning of words in context but not necessarily in isolation. Learning words in context is more effective than learning words in isolation according to studies by F. M. Hamilton, H. Eng, A. Balban, and W. Libby (1, p. 238).

Declarative, interrogative, exclamatory, and imperative sentences were to be understood. The latter two types are extremely limited in Level I, however, so emphasis was placed on declarative and interrogative patterns.

Paragraph structure in Level I is very simple. For the most part a paragraph at this level consists of only three or four simple sentences and is generally a brief description of something concrete or is an answer to a question concerning routine, everyday activities.

Outcome 2 implies that a child must be able to spell accurately in Spanish; he must use accent marks and tildes correctly; he must be able to construct Spanish verbs; and, finally, he must be able to arrange words in their syntactically correct order in declarative and interrogative sentences. As stated before, these specifications are limited entirely to the constructions presented audio-lingually in the Level I television series.

b. Determination of Heuristic Elements

Determination of heuristic elements proceeded concurrently with the development of frames. Some elements were listed empirically prior to beginning the program, but these proved quite inadequate as the program developed. Most of the important elements were discovered through actual experience with the children.

In developing the vocabulary for the program, all of the words and phrases used in the television lessons the first year were arranged alphabetically and a frequency count made. This vocabulary was then checked against Dolch's Basic Sight Vocabulary List and against Keniston's List of Spanish Words and Idioms. The words which were most frequent in the television series and which appeared in Dolch's or Keniston's word lists were selected for use in the first units of the programed instruction; the less common words would appear later in the program. A second alphabetical list was made classifying the various parts of speech separately.

In dealing with grammar, an arbitrary sequence based on experience with traditional text books was set up. As the programer worked with individual children, he continually modified the sequence so that the learner was led to an understanding of that which followed. All grammar was taught functionally, in a normal conversational manner. Grammatical rules, generalizations, or nomenclature were never mentioned. The sequence of grammatical components which finally evolved through this "write and modify" procedure was as follows:

1. General introduction to gender using niño, niña, hermano, hermana.
2. Introduction to gender of inanimate objects such as sombrero, vestido, blusa, casa.
3. Use of possessive nouns (de Roberto, de María).
4. Formation of negative declarative sentences with "no" as the only negative.

   ¿Es Roberto una niña?

   No, no es una niña. Es un niño.
5. Agreement of adjectives with singular nouns ending in "o" or "a".
6. Formation of the plural of nouns ending in a vowel.
7. Indication of first person singular, present indicative, by verb ending "o" and of third person singular, present indicative, by endings "e" and "a".
8. Agreement of singular adjectives ending in letters other than "o" and "a" with singular nouns.
9. Formation of the plural of nouns ending in consonants.
10. Agreement of adjectives with plural nouns.
11. Indication of third person plural, present indicative, by verb endings "en" and "an".
12. Formation of negative questions with "no" as the only negative.

    "¿No juega al beisbol el niño?"
13. Use of the subject pronouns yo, él, ella, ellos, ellas.
14. Indication of first person plural, present indicative, by the verb endings "amos", "emos", and "imos."
15. Position of descriptive adjectives.
16. Use of the idioms "va a", "me gusta", and "le gusta."
17. Employment of the possessive pronouns "mi", "su", "nuestro", "mis", "sus", and "nuestros."

18. Placement of accents on "qué", "quién", "cómo", and "dónde" to form interrogatives.
19. Use of capitalization.
20. Need for definite articles with titles of address.

None of the foregoing components required a specific number of frames; some were presented adequately in twenty frames, while others took as many as sixty. Each component, moreover, marked only the introduction of the particular concept and the basic development of it. The concepts were not presented as entities. On the contrary, every effort was made to constantly reinforce all material previously introduced by frequently presenting it in new contexts and by incorporating the old with the new in gradually more complex structures.

As each new component was introduced, new words were incorporated which lent themselves to the teaching of the particular grammatical concept involved. The words were still selected, however, in relation to the frequency count previously mentioned. Some components had only six new words or phrases to be used in combination with established vocabulary; others had as many as thirty new words.

A useful device for keeping track of vocabulary entry, re-entry, and frequency count was a duplicated alphabetical work sheet on which were printed the first fifty words programed. To this list was added each new word with the number of the component in which it appeared. After each component had been completed and used with individual pupils, a tally was made of the number of times the pupils had been required to write each new word. It was assumed, since there was no evidence on the matter, that each word to be learned should be used actively by the learner at least five times. Although there were a few exceptions, this assumption proved to be generally sound and was used throughout the program.

In addition to developing a sequence of vocabulary and a pattern for teaching grammatical concepts, the programer had to determine the kind of prompts which could be used effectively. The prompts developed are listed below in the order of their effectiveness (with the most effective prompt first).

1. Repetition of a key word:

   Margarita tiene una blusa blanca; la blu_____no es negra.

2. Pictures:

   Hay tres niños y dos _______.

3. Limited choice:

   Los dos hermanos se ___________José y Federico.

4. Parallel construction:

   Roberto es un niño y María es una n_________.

5. Opposites:

   Esta casa es fea, no es bo________.

6. Indicating categories:

   El color de una hoja es _________. (roja, rojo)

7. Drawing a conclusion:

   En la familia Smith hay______personas: la madre y el padre.

8. Use of different tense or number:

   Hay muchas calles en Denver. Una________se llama Downing.

9. Physical arrangement: underlining, italics, different colors, spatial arrangement arrows, etc.

Other elements which proved important are listed below.

1. Words that are spelled alike in both languages but are not related in meaning should be given special attention by placing them in a context which will prevent, if possible, an erroneous interpretation.

2. Use words that are very familiar audio-lingually to teach spelling differences:
   a. Once, doce, trece, catorce, quince for final "e's."
   b. José, rojo, and hijo for "j."
   c. Amarillo and "¿Cómo se llama usted?" for "ll."
   d. "Azul" for "z."
   e. "Pequeño" for que and ñ.
   f. Señor and señorita for ñ.
   g. "Hasta la vista", "Hasta mañana" and hijo for "h."
3. Questions containing "or" should be given careful attention. "Does Mary have a green or yellow dress?" will generally elicit the answer "yes" instead of the color.
4. Call attention to accent marks and tildes. Children will not use them automatically.

c. Development of Individual Frames

To begin the program, forty-one frames requiring multiple choice responses were written on 3 x 5 cards and were used with nine average sixth graders. The programer worked individually with the pupils, explaining to each the purpose of the work to be done and emphasizing that this was not a test but a new way to learn to read and write in Spanish. As a child worked through the sequence of frames, the programer kept a record of the time required for completion, made notes on incorrect responses, and questioned each pupil concerning his responses.

As each child finished, the program was revised by adding or deleting frames; underlining key words; adding pictures, circles, and/or arrows; or by changing the wording. Then the revised set of frames was used with another individual. This process was continued until the nine pupils had worked through the program. By this time, the 41 frames had increased to 107. The time required for completion of the sequence shows the simplification that occurred; the first child had spent thirty minutes on 41 frames while the last child spent only forty minutes on 107 frames.

Many technical faults of the program were discovered immediately and were immediately remedied. This reduced the error-rate somewhat. It soon became apparent, however, that a change in general outlook was necessary for further improvement. It had been assumed that the pupils would know about the same words and grammatical concepts, since all had similar exposure to televised lessons and classroom practice. This proved false. Errors, except those caused by faulty prompts, did not fall into a consistent pattern, indicating a lack of uniformity in the Spanish background of the children. It seemed necessary, therefore, to treat practically every word as a new item to be taught and to keep learning steps between frames quite small.

A change was also made in mode of response. During the first nine trials, multiple choice was the mode employed. During the next nine trials, however, constructed response was used and fewer errors occurred. This change is consistent with findings of Edward B. Fry in a study at the University of Southern California (2). Fry compared constructed and multiple choice response in teaching Spanish vocabulary to English-speaking ninth grade students. He found that, if recall was the learning criterion, constructed response was superior. Because recall as well as recognition is a desired behavioral outcome of the Denver-Stanford project, and because of the lower error rate, the remainder of the program was written entirely with the constructed response mode.

A further improvement was made by adding panels which showed pictures identified by Spanish words. Pupils were encouraged to refer to these panels at any time. Prior to this decision, two other prompting methods were tried. In the first, brief information panels were interspersed with the frames. Each panel was to be studied before continuing on to the next frame and could not be used for further reference. In the second method, the learner studied a panel containing all the difficult vocabulary before beginning the program and he could not later refer back to the panel. These two methods proved ineffective, while allowing the learner to have the panel before him without restriction reduced the error rate considerably. Also, a short quiz given at the end of each trial indicated that more learning occurred when the panel was used continuously.

d. Trials with Larger Groups

Upon completion of thirty-one individual trials the preliminary unit seemed to be relatively error free and educationally sound. The next step was to use the program with a group of children in a regular classroom situation.

The first fifty-three frames, which took the average child twenty-five minutes to complete, were duplicated and stapled in a programed text (panel book) form. The sequence was used with a class of thirteen children, all of whom had viewed the Level I televised Spanish series the year before. Eleven children finished and two did not. The number of errors, given in Table I, gave an overall error rate of .054.

Table 1.

NUMBER OF ERRORS IN A CLASSROOM TRIAL OF THE FIRST 53 FRAMES

NUMBER OF PUPILS	NUMBER OF ERRORS
4	0
3	1
2	2
1	3
1 (Did not finish)	5
1	10
1 (Did not finish)	12

In a quiz administered at the end of the period, nine children made 100%, two made 85%, and two made 50%.

With the programing method now firmly established, arrangements were made to work 25 minutes a week with each of twenty pupils for the remainder of the year. The programer would write from forty to fifty frames before meeting the first child, and revisions were made as the child worked through the frames. The next child worked with the revised program, and further changes were made if needed. In this manner, each individual worked through the program as revised in the preceding trial. A record was kept showing the number of errors, the specific errors, the length of time required to finish each sequence, unit quiz results, and any helpful or interesting comments made by the children.

At the end of each period a brief quiz was given to see whether the child had grasped the intended word meanings and concepts. If the majority of children did not pass the quiz, the sequence of frames was rewritten.

When seventeen of the twenty children had finished working with the first 336 frames, a more extensive test was given. Table 2 shows the results.

Table 2.

RESULTS OF A TEST ON THE FIRST 336 FRAMES

NUMBER OF PUPILS	TEST SCORE
11	100%
3	91%
2	73%
1	64%

After the sixth session, an attempt was made to get opinions from each child concerning the program. All indicated in one way or another that they thought it was a good way to learn; several said it was fun. One girl remarked, "You don't learn any conversational Spanish, but you sure learn to read and write."

Working with twenty children proved to be a slow process. To complete the program in time for use in the 1961-62 school year, the number was reduced from twenty to six. Five children were chosen primarily because they had higher average error rates than the other -- one girl, in particular, nearly always made an error if it were possible. These five children had above average I.Q.'s. The sixth child had a relatively low average error rate but a more nearly average I.Q. (KA-103). The basic method of program building remained unchanged.

The first 336 frames were duplicated and put in panel book form for use with classroom groups. Five classes, representing a wide range of ability and background among pupils, were chosen. Two were from schools in low socio-economic neighborhoods; two were from middle socio-economic areas; and one large class of fifty-seven pupils was from an upper area.

A two part test was composed. The first part utilized pictures as keys to the answers; the child had no opportunity to copy the test item word from the context. This was designed to test accuracy in spelling and in the use of diacritical marks. The second part was written entirely in Spanish and required complete Spanish sentences to be written for answers. Getting a correct answer depended upon an understanding of the context and upon skill in organizing the words into sentences.

The test was administered two weeks prior to the first day of work with the programed text. The classes worked on the program three times a week for thirty-minute periods, and eight periods were allowed for completion of the text. The test was readministered during the ninth class period, five weeks after the original testing.

II. Preliminary Results

Analyses of test results and individual pupil statistics compiled in developing the program follow. These results are preliminary, of course, in that they deal only with pupils participating in the program development phase. The more important and meaningful analysis will involve much larger numbers of pupils using automated materials and will compare their learning with that of pupils instructed by the traditional teacher-directed methods. This material will be the subject of a later report.

In the meantime, the preliminary results indicate that pupils definitely learned from the automated program and show some interesting relationships between time rate, error rate, and amount learned. Group statistics, by school, are given in Table 3.

Table 3.

GROUP STATISTICS RELATED TO USE OF THE FIRST 336 FRAMES

School	Socio-Economic Level of Neighborhood	N	Pre-Test Mean	Post-Test Mean	Gain	Mean IQ	Mean Error Rate	Mean Time Required*
A	Low	29	24.69	68.76	44.07	110.38	.044	159.48
B	Low	28	16.39	27.79	11.40	102.14	.097	141.25
C	Middle	18	33.94	54.78	20.84	112.11	.058	99.07
D	Middle	26	26.23	48.96	22.73	107.35	.078	116.89
E	High	57	30.02	53.00	22.98	112.54	.036	137.40
Total		158	26.45	50.96	24.51	109.40	.058	134.39

*Number of minutes required to complete 336 frames.

Table 4 shows tests of the significance of differences between pre-post-test means for all schools combined and for school B (the group with the smallest gain).

Table 4.

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS ON TESTS COVERING FIRST 336 FRAMES

School	Difference	σ_{D_M}	Critical Ratio	Probability
B	11.40	1.69	6.74	$< .001$
Total	24.51	1.27	19.30	$< .001$

With both probabilities less than .001, there can be no doubt that all of the pupils learned from the program. The Pearson product-moment correlation coefficient between pre- and post-tests was .717, incidentally, which indicates a rather even rate of learning among most pupils.

Two groups, however, showed large variations, and both were from neighborhoods classified low in socio-economic status. The group from school A had by far the largest gain, and that from school B had the smallest gain. In an attempt to determine causal factors, rank correlations (rho's) based on group standings were computed between the variables. Table 5 gives these rho's. In this computation, the group with the smallest error rate was ranked first; for other variables, the group with the highest mean was given the highest rank.

Table 5.

RANK CORRELATIONS
BETWEEN PAIRS OF VARIABLES

VARIABLES	RHO
gain x IQ	.50
gain x error rate	.80
gain x time rate	.40
IQ x error rate	.50
IQ x time rate	-.30
error rate x time rate	.10

The IQ relationships are about as expected. Higher IQ pupils showed greater gain, made fewer errors, and took less time than others. The very high relationship between gain and error rate was reassuring since the minimum error philosophy of programing had been adopted. Somewhat surprising, however, was the lack of relationship between error rate and time rate, especially since time rate also correlates with gain. Making fewer errors and taking more time both increase learning, but they seem to work independently.

These relationships account only in part for the great difference in gain between groups A and B, and they leave some perplexing questions unanswered. Why, for example, should group B's error rate be more than twice that of group A? Why should group A take more time to complete the program than group B? With the negative relationship between IQ and time and the higher average IQ of group A, the A's would be expected to take less time.

Evidently group A was more interested in the program and had greater desire to learn from it. And, since the program writer explained the text and instructed all groups in the use of it, group A's greater motivation would appear to result from common experiences before and/or during their use of the book. This suggests that the classroom teacher had a great effect on the amount pupils learned even when using automated instruction materials.

The second 336 frames had been tested on individuals and printed in panel book form by the time the post-test was taken, and the groups from Schools A, B, and D continued with the program. Other activities prevented the continuation of schools C and E.

Again pre- and post-tests were given, and results, shown in Table 6, were consistent with those of the first set of frames. The Pearson product-moment correlation coefficient between pre- and post-tests was .717 -- the same as that for the first tests.

Table 6.

GROUP STATISTICS RELATED TO
USE OF THE SECOND 336 FRAMES

School	N	Pre-Test Mean	Post-Test Mean	Gain	Mean Error Rate
A	29	21.90	77.76	55.86	.020
B	28	6.48	21.30	14.82	.064
D	26	13.17	31.33	18.16	.057
Total	83	13.49	43.84	30.35	.046

Again, a significant amount was learned by all pupils, as the comparison of means in Table 7 shows.

Table 7.

SIGNIFICANCE OF
DIFFERENCES BETWEEN MEANS
ON TESTS COVERING SECOND 336 FRAMES

School	Difference	σ_{D_M}	Critical Ratio	Probability
B	14.82	3.47	4.27	$< .001$
Total	30.35	3.30	9.21	$< .001$

As stated before, these results are preliminary. They show that pupils definitely learn from the automated instruction program, and this is important. But the real value of the program cannot be assessed until it is compared to more conventional teacher-directed instruction.

The results also indicate that factors other than ability and outside of the program itself affect learning. The instruction is automatic in the sense that the classroom teacher is not directly involved, but the amount learned from the program is not automatic. Determining the causative factors could be extremely important, and further research is obviously needed.

Automated instruction materials will be used in the Denver-Stanford project the following two school years with thousands of sixth grade pupils. Initially, efforts will be made to determine the value of automated instruction in terms of learning for the type of subject matter and level of pupil involved, and secondly to isolate some of the factors related to the amount learned.

This report describes only the development of the first 672 frames -- about one-third of the total anticipated program. The developmental process as it finally emerged seems sound and is being employed in building the remainder of the program. It is hoped that the Denver experiences reported here will be useful to other programers facing similar problems.

REFERENCES

(1) Buchanan, M. A. and E. D. MacPhee. "Factors Affecting Learning." In Twentieth Century Modern Language Teaching. Maxim Newmark, ed. New York: The Philosophical Library, 1948. pp. 237-247.

(2) Fry, Edward B. "A Study of Teaching Machine Response Modes." In Teaching Machines and Programmed Learning. A. A. Lumsdaine and Robert Glaser, eds. Washington, D.C.: National Education Association, 1960. pp. 469-474.

(3) Fry, Edward B., Glenn L. Bryan, and Joseph W. Rigney. Teaching Machines: An Annotated Bibliography. Special Edition of Audio Visual Communication Review. 1960, 8, No. 2.80 pp.

(4) Galanter, Eugene, ed. Automatic Teaching: The State of the Art. New York: John Wiley and Sons, Inc., 1959. 198 pp.

(5) Lumsdaine, A. A. and Robert Glaser, eds. Teaching Machines and Programmed Learning. Washington, D.C.: National Education Association, 1960. 724 pp.

THE AUTHORS

Delbert Barcus is a language specialist on the Denver-Stanford project staff and is the developer of the programed learning materials discussed herein. He taught Spanish in the Denver Public Schools for seventeen years before joining the project staff.

John L. Hayman, Jr., is director of the Denver-Stanford project for Stanford University and is specifically responsible for research methodology and analysis in the project.

James T. Johnson, Jr., is director of the Denver-Stanford project for the Denver Public Schools and serves as overall coordinator of the project.

ED035269

RESULTS OF THE FIRST YEAR'S RESEARCH IN THE DENVER-STANFORD PROJECT

- JOHN L. HAYMAN, JR.
- JAMES T. JOHNSON, JR.

EM007066

Denver Public Schools • Stanford University

RESEARCH ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

ED035269

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

RESULTS OF THE FIRST YEAR'S RESEARCH IN THE DENVER-STANFORD PROJECT

by

John L. Hayman, Jr.
James T. Johnson, Jr.

DENVER-STANFORD PROJECT
ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

* * * * * *

School District Number One
City and County of Denver
Denver, Colorado

Institute for Communication Research
Stanford University
Stanford, California

Report Number 5
January 1962

SUMMARY

The Problem

A large number of studies has established conclusively that instructional television is an effective teaching device. In the great majority of comparisons, students have learned at least as much through television as they have through the more traditional, face-to-face classroom situation. Almost all of these studies, however, have employed television as a unique teaching device and have not considered the context in which it was used. This type of research was necessary to establish the usefulness of television in education, but it does little to define TV's ultimate potential.

Experts in the field have generally agreed that the total learning process, in which television plays a role, must now be considered. The Denver-Stanford project was begun in 1960 with this aim, with the study of TV presentation methods and classroom and home activities surrounding television. Using elementary school Spanish in a field research setting, the project is attempting to increase the effectiveness of instructional television through systematic variation of its context.

Results

The first full school year of research in the project was 1960-61. During that year, the research was limited to the fifth grade level, and some 6,000 pupils took part. Six research groups were defined for the first semester. All pupils viewed a 15-minute television lesson in school on Mondays, Wednesdays, and Fridays. The groups differed according to activities in addition to these lessons. The groups and their activities were: (1) no additional activity, (2) a second viewing of the lesson in the home at night, (3) 15 minutes of dialogue practice in the classroom after the lesson, (4) 15 minutes of structure practice in the classroom, (5) 15 minutes of eclectic practice in the classroom, and (6) a second viewing of the lesson at night with parents and other parent help at home. Mid-term and final tests were administered during the semester. Group two performed significantly better than group one on these tests, group six performed significantly better than group two, and each of the classroom practice groups performed significantly better than group six. Among those with classroom practice, the eclectic group performed better than the structure group which, in turn, performed better than the dialogue group; the only statistically significant difference, however, was that between the eclectic and dialogue groups.

These results were used to restructure the experimental design for the second semester. Group one was eliminated, since it showed little promise, and the other five procedures were continued. In addition, six groups combining classroom and home activities were established. For three of them, a second viewing at home was combined respectively with dialogue, structure, and eclectic classroom practices; and for the other three, a second viewing with parent help was combined respectively with dialogue, structure, and eclectic classroom practices. In the test given at the end of the semester, relative performance among the five original practice groups was the same as it had been on first semester tests. The combination of evening viewing with classroom practice seemed to make little difference; i.e., these groups performed similarly to those with classroom practice alone. The combination of evening viewing and parent help with classroom practice made a significant difference, however, and groups in this category performed better than any others. Further study of evening activities showed that pupils who merely viewed with their parents learned more, irrespective of additional activities, than those who viewed alone. The value of other parent-pupil activities appeared to increase as the year progressed.

Covariance analysis was the main statistical technique employed to arrive at the above results. The dependent variable in each case was performance on listening comprehension tests developed by project personnel. A test of oral ability was also used at the conclusion of the second semester.

TABLE OF CONTENTS

PART PAGE

I. BACKGROUND OF THE PROJECT 1

II. RESEARCH DESIGN FOR THE 1960-61 SCHOOL YEAR 5

- Research Groups for the First Semester 5
- Assignment of Pupils and Teachers to Groups 8
- Teacher Training 9
- Information Collected 10
- Research Groups for the Second Semester 12

III. FINDINGS 13

- The First Semester 13
- The Second Semester 19
- Causal Factors Related to Double Viewing and Parent Help . . 30
- Summary of Findings 34

IV. DISCUSSION OF FINDINGS AND OF FUTURE PLANS 35

RESULTS OF THE FIRST YEAR'S RESEARCH IN THE DENVER-STANFORD PROJECT

by

John L. Hayman, Jr.
James T. Johnson, Jr.

The Denver Public Schools and Stanford University's Institute for Communication Research are currently engaged in a joint research project on the context of instructional television. The purpose of the project is to learn how instructional television can best fit into the total teaching situation. A substantial amount of research has established that television is a very effective teaching medium. Ways of combining it with other educational activities must now be considered, and the Denver-Stanford project is a beginning effort in this direction. Kenneth E. Oberholtzer is principal investigator for the Denver Public Schools and Wilbur Schramm is principal investigator for Stanford University. This is one of a number of project progress reports.

I. BACKGROUND OF THE PROJECT

Some 400 scientifically designed and statistically treated studies on educational television (ETV) have now been completed, and they have dealt with a wide variety of subjects at all levels of instruction -- college, high school, and grade school. In the great majority of these studies, (84 per cent), students instructed by television have learned as much as or more than students with conventional classroom instruction. As Schramm states in his extensive review of ETV research, "There can no longer be any doubt that students learn efficiently from instructional television" (Schramm, 1961, p. 1).

These results are of unquestionable importance. However, Runyon and Kanner's 1956 observation, that "most (ETV) researchers have been content to ask the general question, 'How does television instruction compare in effectiveness to regular instruction?'" is still true (Runyon and Kanner, 1956, p. 84). The answer to this question would be sufficient perhaps, if ETV were always to handle the complete teaching load for a subject, i.e., if educators were faced with choosing either televised instruction or classroom instruction.

This is not the choice faced, however. It is possible to use ETV in a variety of combinations with classroom and home activities, and television could well be most effective in such combinations. Indeed, the few early studies considering such usage suggest that this is the case, that using television in conjunction with teacher-directed classroom activities increases both learning and motivation of pupils and makes television more acceptable to teachers. But we still know little about what these activities should be.

This situation leads Schramm to conclude:

> . . . the most important research on instructional television, now, . . . is research on the total process of which television is a part. When Carpenter talks about television research as system research (Carpenter, 1960), he is, in effect, asking how television fits into the learning experience. How can it be used best, and for what? How can it best be combined with other experiences, to make learning a given subject most efficient (Schramm, 1961, p. 28)?

The Denver-Stanford project, which has the official title, "Four Years of Research on the Context of Instructional Television," is designed to consider some of the problems which now seem most important in ETV. In it the whole teaching situation of which television is a part is studied, with as many elements varied as can be handled. The attempt is made to see where television fits into the pattern and what combinations of activities and facilities can be put together with it to increase learning efficiency.

The project was conceived in 1959 when the Denver Public Schools and the Stanford Institute for Communication Research were brought together by the National Educational Television and Radio Center (NETRC), whose principal officials were concerned with the need for research on the context of ETV. A joint project was suggested so that the facilities and interests of both a large city school system and a research institute could be brought to bear on the problem.

Foreign language in the elementary school was selected as the subject matter to be taught, with Spanish the specific language. There were several reasons for this choice. For one thing, a shortage of elementary school classroom teachers qualified to teach foreign language makes the use of television a necessity in this area. For another, foreign language is an addition to the elementary school curriculum which seems to be called for, and which has already begun in many school systems. Furthermore, it is an innovation which has not been frozen by years of use and thus lends itself well to experimentation.

A proposal was submitted to the United States Office of Education in June, 1959, for a grant under Title VII of the National Defense Education Act, and this was approved a few months later. A joint research team from the two organizations was then created to be directly responsible for the tasks required. The Institute for Communication Research assigned research personnel to this team and the Denver Public Schools assigned specialists in education, language, and television teaching and production. In addition, Denver made available the facilities of KRMA-TV, an educational station which it operates.

Actual work on the project began in January, 1960, with these general hypotheses to be tested:

1. Instructional television of a certain kind, reinforced by certain classroom activities, makes for more learning, ceteris paribus, than without this reinforcement.
2. Instructional television of a certain kind, reinforced by certain home activities, makes for more learning, ceteris paribus, than without this reinforcement.
3. Instructional television of a certain kind, reinforced by certain home and classroom activities, makes for more learning, ceteris paribus, than without these reinforcements or with either one alone.
4. Certain kinds of instructional television, combined with certain kinds of home and classroom reinforcements, make for more learning, ceteris paribus, than other kinds and combinations.

These imply systematic variation of what pupils see on the television screen; of classroom activities, including pupil use of tape recorders, record players, automated instruction materials, and the like, in addition to teacher-conducted instruction; and of home activities, including those with and without parent participation.

The first months of the project were used mainly for orientation of project personnel and for planning and preparing for the 1960-61 school year. A preliminary experiment was conducted in May, 1960, however. Some 550 fifth grade pupils in their first year of Spanish instruction were used to compare three methods of television presentation. The methods were (1) lecture, (2) modified lecture with elicited student participation in the classroom, and (3) modified lecture with instructor-student interaction in the studio and student participation in the classroom. Covariance analysis of overall between group differences gave a frequency ratio of

2.054 which, with n_1 of 2 and n_2 of 506, is significant at approximately the .075 level. Individual group comparisons showed that method three produced the highest mean and method two the lowest, and the difference between groups three and two was significant at the .059 level. Neither of the other differences was significant, though that between groups one and two (critical ratio of 1.181) was relatively larger than that between groups three and one (critical ratio of 0.433). A replication of this experiment is planned (for a complete report, see Hayman, 1961).

The results of the preliminary experiment led to the choice of presentation method three for the television lessons. Thus all pupils in the project have the same television experience: they are led by the television teacher to practice correct responses during the program and they see pupils from their own grade interact with the teacher. To assure that the television lessons are of high quality and incorporate the desired characteristics, and to make certain that the television experience of one group can be repeated later for another group, all lessons are recorded on video-tape prior to the telecasting.

II. RESEARCH DESIGN FOR THE 1960-61 SCHOOL YEAR

a. Research Groups for the First Semester

The first year's research was designed to be conducted at the fifth grade level in two stages. In the first semester, a number of single procedures would be compared. The most promising of these would then be tried in various combinations during the second semester.

Foreign language was made a required subject for all fifth graders in the Denver Public Schools at the beginning of the 1960-61 school year, and some 6,500 pupils chose Spanish (French was the other possibility) and were automatically in the project. Every pupil viewed the 15-minute televised lesson on Monday, Wednesday, and Friday of each week, so the television experience was the same for each. The main effects to be compared, then, were activities in addition to classroom television. In other words, the context of the television was changed for different groups.

For the first semester, the research groups and activities in addition to viewing the in-school telecast were:

1. No additional activity. Pupils in this group saw the telecasts in school and did nothing else. Their performance served as a base against which to compare the effectiveness of other activities.
2. A second viewing of the television lesson. Each television lesson was repeated at night on KRMA-TV, and pupils in this group viewed it in their homes. This was a repetition for them since they had seen it earlier in school.
3. Dialogue practice in the classroom. Pupils in this group took part in dialogue practice conducted by their classroom teacher. Each practice session lasted 15 minutes and occurred immediately after the television lesson.
4. Structure practice in the classroom. This group was similar to group 3 except that structure practice was used in the classroom.
5. Eclectic practice in the classroom. This group was similar to groups 3 and 4 except that eclectic practice was used in the classroom.
6. A second viewing of the lesson and parent help. Pupils in this group viewed the television lesson at night with their parents and had other types of parent help at home. This group was the same as group 2 except for parent help.

The activity followed by group two was designed to test the value of repetition, which has often proved a powerful variable in education.

Group six was set up to determine if direct parent aid would be of value to pupils in this situation. Parents were provided with a phonograph record especially designed to complement the television lessons and with a study guide which suggested a number of home activities. Parent participation in the project is described fully in a previous progress report (Hayman and Johnson, 1961).

The teacher-conducted classroom activities of groups three, four, and five are defined as follows:

A dialogue drill consists of a short, meaningful conversation between two people. The pupil learns the part of one person in the conversation with the teacher taking the other part. Then the roles are reversed, with the pupil taking the other part. Finally, two pupils use the drill to converse. Dialogue practice is said to be valuable in language learning because:

> It involves a natural and exclusive use of the audio-lingual skills. . . . All that is learned is meaningful, and what is learned in one part of a dialogue often makes meaning clear in another. From the start the student learns to address people directly and to use the first person singular -- items of basic importance in communication. If the materials are appropriate, the learner finds a personal interest in what he is saying and a possible use far beyond the classroom for the expressions which he masters (Brooks, 1960, p. 141).

Structure drill involves repetition of a language pattern with small and consistent changes at each repetition. The pattern usually consists of a single sentence, and any word or phrase in the sentence may be changed. The following is an example in English with the verb varied: I see an apple; I buy an apple; I eat an apple; I like an apple; etc. Structure practice, as opposed to dialogue, is not meaningful communication *per se*.

> It is to communication what playing scales and arpeggios is to music: exercise in structural dexterity undertaken solely for the sake of practice, in order that performance may become habitual and automatic -- as it must be when the mind concentrates on the message rather than on the phenomena that convey it. Structure practice capitalizes on the mind's capacity to perceive identity of structure where there is a difference in content, and its quickness to learn by analogy (Brooks, 1960, p. 142).

Though dialogue drill and structure drill are both valuable in language learning, neither would ordinarily be used as a single teaching method; rather, some combination would be used. They were, in fact, used in combination with television instruction in the project. A cogent question was whether the normal type of classroom activity rather than an isolated method would be more effective with ETV. To answer this, the eclectic practice was used. This practice combined dialogue and structure drill with other items such as narratives, songs, and games.

b. Assignment of Pupils and Teachers to Groups

Research groups three, four, and five required reasonably well qualified Spanish teachers since they involved active classroom instruction. Teachers in groups one, two, and six, on the other hand, had a passive classroom role and did not have to be as well qualified in Spanish. Therefore, all teachers in the project were classified according to their qualification for Spanish instruction. Prior training and teaching experience were the classification criteria.

Qualified teachers were assigned randomly to groups three, four, and five, and less qualified teachers were assigned randomly to groups one, two, and six. This assignment procedure assured similarity of teaching ability among the three groups with classroom instruction and among those who had only television in the classroom.

No attempt was made to control pupil assignment from the research project office. Over 6,000 pupils in some 90 elementary schools were involved, and scheduling problems in the individual schools made central control impossible. This meant that similarity of learning potential among groups was not assured. With one exception, however, no biasing factor was involved since pupils were assigned to classes before teachers were assigned to research groups. The exception was group six. Classes were not placed in group six unless a majority of parents in the class volunteered to participate, and this selection factor could create differences. (See Hayman and Johnson, 1961, pp. 9-12, for its actual effect.)

c. Teacher Training

For a valid comparison of the dialogue, structure, and eclectic practices, it was necessary for teachers in groups three, four, and five to follow the exact procedure to which they had been assigned. This, in turn, depended on their understanding of the procedure and of the purpose of the project.

A program of teacher training was established under the direction of the Denver Public Schools' Supervisor of Foreign Languages. One phase of this program consisted of summer workshops for which college credit was granted. Fifty teachers, over one fourth of those in the project, attended the workshop in 1960, just prior to the first year of research described herein.

The second phase of the teacher training program consisted of instructional sessions held during the school year. The first of these lasted a half day and occurred before any classes had met. Teachers were told of the background of the project and of the questions for which answers were sought. Their essential role in securing valid results was emphasized. Then instruction in individual procedures was given. Finally, each group received guides which further explained the procedure to be followed and provided classroom drills.

Five other training sessions were held during the school year. Two of these were televised, two were conducted in individual schools, and one was a general meeting at the beginning of the second semester. At each session, teachers concerned with a specific procedure were given instruction in that procedure. The general meeting included a discussion of progress to date, problems encountered, and future plans.

In addition to the formal sessions, staff personnel met with individual teachers during the year to discuss procedures and help with individual problems. These meetings were scheduled on a random basis within each research group to assure that teachers in each group had the same amount of individual help.

As stated above, the primary purpose of the teacher training was to make certain that assigned procedures were followed properly. A second purpose was to increase the proficiency of all teachers in the project, including those not considered qualified for an active role at the beginning of the year. All teachers were asked to participate during the television lessons, i.e., to follow the TV instructor closely and to respond when response was requested. With television instruction, the in-service training sessions, and the summer workshops, it was hoped that every teacher could handle at least the beginning level of Spanish during the following school year (1961-62).

Teacher training will be the subject of a project progress report to be published shortly and will be more fully discussed there.

d. Information Collected

The exclusive use of audio-lingual teaching methodology during the first year of research meant that only listening and speaking skills would be evaluated. This simplified measurement in the sense that it limited the number of instruments needed, but it complicated it in that pupils could neither read nor write in Spanish and translation was strictly avoided. The lack of suitable measuring instruments necessitated development of such instruments by the project staff. Three listening comprehension tests were developed and administered via television during the school year (Andrade, Hayman, and Johnson, 1961). Pupils responded by marking choices (mostly of pictures) on a standard answer sheet, and scoring was handled at the project office. One test was given about two thirds of the way through the first semester to allow restructuring of research groups at mid-year, and the other two tests were finals for the first and second semesters.

Measurement of speaking ability is considerably more difficult. An actual speaking performance of the pupil must be evaluated, and the fact that more than 6,000 pupils were involved made scoring by a few staff members impossible. Scoring in this situation is difficult and subjective at best, and reliability is certain to be quite low unless well-trained experts are used. Considerable effort went into development of an oral test, which was tried on a limited basis at the end of the first semester. Although this test, after certain revisions, seemed satisfactory in terms of validity and reliability, staff members felt that it could not be administered and scored satisfactorily on a broad basis. The solution for 1960-61 was to choose a random sample of 40 pupils each from research groups two through six and test these pupils the last week of the second semester. The ultimate goal is to develop oral tests which can be administered and reliably scored by the average classroom teacher. A later progress report will be concerned with measurement of speaking skills.

These test results constituted the criterion or dependent variables for the first year of research. As mentioned previously, however, the pupil assignment method did not assure similarity of learning ability among the research groups. The obvious statistical technique for this type of situation is covariance analysis, which controls for group differences. To use covariance, measurements on key variables related to the dependent variable must be available.

A number of scores in addition to those of the criterion tests were therefore collected for each pupil. These included: IQ, age, grade point average in academic subjects (GPA), sex, occupation of the head of the household (as an indication of socio-economic status), and Stanford Achievement Test scores in paragraph meaning (PM), word meaning (WM), spelling (SP), language (L), arithmetic comprehension (AC), and arithmetic fundamentals (AF). Pupils in groups two and six were given calendars on which to indicate the number of evening telecasts they watched, and group six parents were asked to indicate the number of evening telecasts they watched and to show the number of minutes they practiced Spanish with their child each week.

e. Research Groups for the Second Semester

The first listening comprehension test was evaluated in time to serve as a basis for restructuring the research design for the second semester. This test showed conclusively that the practice followed by group one (a single viewing of the TV lessons) was not as effective as the other procedures, and this practice was eliminated. Since the other five practices all showed promise, they were retained. In addition, six combination groups were established. These consisted of combining a second viewing with dialogue, structure, and eclectic classroom practice and of combining parent help with dialogue, structure, and eclectic classroom practice.

As in the first semester, teachers were assigned to groups and their pupils automatically went with them. A few teachers who had been in groups one and two desired a more active role, and they were assigned randomly to the three classroom procedures. Each classroom procedure group was in turn randomly divided into two parts, and one part was given the second viewing in the evening. For the most part, then, teachers in the classroom practice groups and the classroom practice plus evening viewing groups had had an active teaching role the first semester.

Teachers for the parent help-classroom practice combinations had to be drawn from those who had been in the parent help groups the first semester. This was necessary because teachers could not be moved to a different class at mid-year and because no new parent participants were to be added without benefit of the first semester. Therefore, teachers who were with parent-help pupils the first semester were randomly divided into four groups. One of these continued as a parent help group with no classroom practice and the other three were assigned to classroom practice -- one each to dialogue, structure, and eclectic. This assignment procedure meant that teachers in the classroom practice-parent help groups were on the average not as well qualified to teach Spanish as those in the other classroom practice groups.

III. FINDINGS

a. The First Semester

The first listening comprehension test was administered in December, about two thirds of the way through the fall semester. Means for the six research groups (identified on page 6) on the 50-item test and on the six variables chosen for control purposes are given in table 1.

Table 1

GROUP MEANS
ON THE DEPENDENT AND CONTROL VARIABLES

Variable	Group						
	1	2	3	4	5	6	Total
Test	26.813	29.391	33.398	33.544	33.816	32.694	31.609
IQ	99.844	100.018	99.380	100.960	99.482	100.926	100.100
PM*	51.735	52.650	50.793	51.827	52.629	54.901	52.422
WM*	53.003	52.439	51.425	53.095	52.779	54.252	52.832
SP*	50.636	50.643	50.864	50.129	50.129	50.575	50.799
L*	52.595	50.952	58.493	54.956	50.010	57.531	54.256
GPA*	2.591	2.614	2.578	2.588	2.576	2.662	2.602

*Stanford Achievement Scores (PM, WM, SP, L) are in terms of grade placement times ten. Grade point average (GPA) was computed as follows: D = 1, C = 2, B = 3, and A = 4.

The control variables were selected on the basis of experience gained in the May preliminary experiment and of a series of multiple correlations between the dependent variable and different combinations of potential control variables. The variables not used were rejected even though they had a low positive correlation with test score. They could add nothing to the analysis because all the variance they had in common with test score was accounted for by the control variables listed in table 1. In fact, further study showed that WM, SP, and L could also be eliminated without changing the results of the analysis; therefore, IQ, PM, and GPA were the only control variables used in other statistical work for the remainder of the year.

Pearson product moment correlations between pairs of variables are shown in table 2.

Table 2

CORRELATIONS BETWEEN
PAIRS OF VARIABLES

Variable	Variable					
	IQ	PM	WM	SP	L	GPA
Test	.392	.395	.387	.372	.364	.371
IQ		.701	.729	.704	.615	.683
PM			.826	.710	.629	.683
WM				.755	.645	.672
SP					.633	.699
L						.601

A one-dimensional covariance analysis was run to test the overall differences between groups. As table 3 indicates, this analysis gave an F-ratio of 67.359 which, with n_1 of 5 and n_2 of 1752, had a probability less than .001. For this analysis, a random sample of 294 pupils was drawn from each of the six research groups, giving a total N of 1764.

Table 3

COVARIANCE ANALYSIS OF
THE FIRST SEMESTER MID-TERM

Source of Variation	Residual		
	DF	Sums of Squares	Mean Square
Total Groups	1757	73,319.270	
Within Groups	1752	61,497.260	35.101
Difference (Between Groups)	5	11,822.010	2,364.402

$$F = \frac{2,364.402}{35.101} = 67.359$$

$n_1 = 5$ $n_2 = 1752$

$P < .001$

The covariance analysis leaves no doubt that real differences among groups were found. To compare the individual groups, mean scores were adjusted by the covariance technique, and critical ratios of differences between group means were computed. The standard deviations used were derived from the group variances given in table 4. Table 4 also shows the adjusted group means. Table 5 gives the obtained critical ratios and their probabilities.

Table 4

ADJUSTED GROUP MEANS AND VARIANCES
FOR THE FIRST SEMESTER MID-TERM

	Research Group						
	1	2	3	4	5	6	Total
Group Mean	26.909	29.448	33.499	33.519	33.914	32.341	31.609
Group Variance	54.549	60.009	55.935	45.937	48.726	45.937	51.880

Table 5

SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS
FOR THE FIRST SEMESTER MID-TERM

Groups	Difference	σ_{DM}	CR	P
2 - 1	2.539	.624	4.068	<.001
6 - 2	2.893	.600	4.821	<.001
3 - 6	1.158	.588	1.969	=.049
4 - 3	0.020	.588	0.034	>.400
5 - 3	0.415	.596	0.696	>.200
5 - 4	0.395	.567	0.696	>.200

Table 5 shows that a second viewing at night (2) was significantly superior to merely watching the television lessons during school (1), that a second viewing at night plus parent help (6) was significantly superior to the first two procedures, and that having classroom instruction (3, 4, and 5) was significantly superior to anything else. No significant differences occurred among the three classroom instruction groups, although the eclectic group had the highest mean and the structure practice group was second.

The second listening comprehension test was given the last week of the first semester. It also contained 50 items, 23 of which were repeated from the first test. The same subjects were used in the analysis so far as possible. Absences during the test administration necessitated replacing 309, however. Replacements were drawn randomly from the different groups. Control variable means and correlations between variables remained about the same as indicated in tables 1 and 2. The correlation between the first two tests was .735.

Table 6 gives unadjusted group means for the first semester final.

Table 6

UNADJUSTED GROUP MEANS
FOR THE FIRST SEMESTER FINAL

				Group			
	1	2	3	4	5	6	Total
Mean	30.491	33.087	37.497	38.484	38.797	35.535	35.644

Results of the one-dimensional covariance analysis of the first semester final are shown in table 7. In this analysis, the F-ratio was 61.014, and the probability again was far less than .001, indicating real differences among the groups.

Again group means were adjusted and differences between them tested for significance. Variances and adjusted means are given in table 8, and table 9 shows the significance tests.

Table 7

COVARIANCE ANALYSIS OF THE FIRST SEMESTER FINAL

Source of Variation	Residual		
	DF	Sums of Squares	Mean Square
Total Groups	1725	114,224.232	
Within Groups	1720	97,016.748	56.405
Difference (Between Groups)	5	17,207.484	3,441.496

$$F = \frac{3,441.496}{56.405} = 61.014$$

$n_1 = 5$ $n_2 = 1720$

$P < .001$

Table 8

ADJUSTED GROUP MEANS AND VARIANCES FOR THE FIRST SEMESTER FINAL

	Research Group						
	1	2	3	4	5	6	Total
Group Mean	30.455	33.020	37.895	38.646	39.007	34.868	35.644
Group Variance	85.340	80.222	70.259	66.799	58.821	62.836	79.819

Table 9

SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS
FOR THE FIRST SEMESTER FINAL

Groups	Difference	σ_{D_M}	CR	P
2 - 1	2.565	.756	3.392	<.001
6 - 2	1.848	.704	2.625	=.008
3 - 6	3.027	.680	4.451	<.001
4 - 3	0.751	.691	1.086	>.100
5 - 3	1.112	.669	1.662	=.048* =.097*
5 - 4	0.361	.659	0.547	>.200

*P = .048 on a one-tailed test
P = .097 on a two-tailed test

A comparison of means in tables 4 and 8 and of differences in tables 5 and 9 shows a high degree of consistence between the two tests administered during the first semester. The rank order of groups was exactly the same, and differences between the groups were quite similar. The probabilities were similar, in fact, with two exceptions. First, group five (eclectic classroom practice) gained in relation to groups three and four so that it was significantly superior to three on a one-tailed test. Second, group six slipped somewhat relative to the others; it was closer to group two, although still superior at the .008 level.

b. The Second Semester

The second semester listening comprehension test was administered via TV in May. This 50 item test covered content of the second semester only and did not include any items used in previous tests. Of the 2,002 pupils included in the analysis, 1400 were used in analysis of the first semester final and 602 were new additions. The research groups and the number of pupils in the analysis from each group are shown in table 10.

Table 10

GROUPS AND NUMBERS OF SUBJECTS IN THE ANALYSIS OF THE SECOND SEMESTER FINAL

Group Number	(First Semester Identification)	Practice Followed: TV Viewing in the Classroom Plus	Number of Subjects
1	(2)	A Second TV Viewing at Night	91
2	(3)	Dialogue Practice in the Classroom	166
3	(4)	Structure Practice in the Classroom	210
4	(5)	Eclectic Practice in the Classroom	161
5	(6)	A Second TV Viewing at Night and Parent Help	116
6	(2-3)	A Second TV Viewing at Night plus Dialogue Practice in the Classroom	272
7	(2-4)	A Second TV Viewing at Night plus Structure Practice in the Classroom	274
8	(2-5)	A Second TV Viewing at Night plus Eclectic Practice in the Classroom	291
9	(6-3)	A Second TV Viewing at Night and Parent Help plus Dialogue Practice in the Classroom	138
10	(6-4)	A Second TV Viewing at Night and Parent Help plus Structure Practice in the Classroom	149
11	(6-5)	A Second TV Viewing at Night and Parent Help plus Eclectic Practice in the Classroom	134
		TOTAL	2002

The May test was examined in two ways. In one analysis, the first semester final was used as a control variable, so that only the amount learned during the second semester was reflected. This procedure was appropriate for examining those practices which were instituted at the start of the second semester. In the second analysis the first semester test was not used as a control variable, and the amount learned during the whole year was reflected. This procedure was used in examining practices followed throughout the year. Groups one through five followed the same practice both semesters, and groups six through eleven had a new practice introduced in the second semester.

A one-dimensional covariance analysis, with IQ, grade point average, and Stanford achievement score in word meaning used as control variables, was run for each situation. Results of these analyses are given in table 11.

Table 11

ONE-DIMENSIONAL COVARIANCE ANALYSES
WITH AND WITHOUT THE JANUARY TEST USED AS A CONTROL VARIABLE

With the January Test				Without the January Test			
Source of Variation	Residual			Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square		Degrees of Freedom	Sum of Squares	Mean Square
Total	1997	25,509.556		Total	1998	84,469.610	
Within Groups	1987	23,975.051	12.066	Within Groups	1988	78,502.071	39.488
Between Groups	10	1,534.505	153.451	Between Groups	10	5,967.539	596.754
$F = \frac{153.451}{12.066} = 12.718^*$ $P < .001$				$F = \frac{596.754}{39.488} = 15.112^*$ $P < .001$			

*F of 7.64 required for .001 level of significance.

With or without the first semester final (January test) as a control variable, the overall results were highly significant, and examination of group means was in order. Unadjusted group means, means adjusted for the first semester final, and group variances are given in table 12.

Table 12

UNADJUSTED MEANS, MEANS ADJUSTED FOR THE JANUARY TEST, AND GROUP VARIANCES

Group	Unadjusted Mean	Mean Adjusted for January Test	Variance
1	27.901	32.057	57.520
2	34.120	35.251	58.327
3	35.519	35.709	57.556
4	36.062	36.075	39.143
5	34.802	33.001	40.839
6	35.066	34.830	53.774
7	35.575	35.472	50.790
8	35.572	35.503	51.598
9	37.116	36.464	34.492
10	36.671	36.218	35.004
11	36.627	35.671	39.080
Total	35.283	35.283	51.529

Again the outcome is consistent with previous findings. However, there was one rather surprising result. The classroom practice plus evening viewing had a lower mean in each case than classroom practice alone; i.e., dialogue classroom practice plus evening viewing (6) was lower than dialogue classroom practice alone (2); structure practice plus evening viewing (7) was lower than structure practice alone (3), and eclectic practice plus evening viewing (8) was lower than eclectic practice alone (4). One would expect the second viewing to increase learning. Significance tests in table 13 show that each difference had a probability greater than .35, however, so these differences must be considered chance fluctuations and the null hypothesis -- that evening viewing made no difference when combined with classroom practice -- accepted.

Table 13

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS
WHERE ONLY SECOND SEMESTER LEARNING IS REFLECTED

Groups Compared	Difference	σ_{D_M}	Critical Ratio	Probability
2 - 6	.421	.741	.568	> .50
3 - 7	.237	.678	.350	> .50
4 - 8	.572	.648	.883	> .35
9 - 2	1.213	.775	1.565	= .12
10 - 3	.509	.713	.714	> .40
4 - 11	.404	.731	.553	> .50
11 - 5	2.670	.802	3.329	< .001
6 - 1	2.773	.991	2.798	< .01
5 - 1	.944	.817	1.155	= .13*
2 - 5	2.250	.838	2.685	< .008
3 - 2	.458	.790	.580	> .30
4 - 2	.824	.771	1.069	= .14*
7 - 6	.642	.619	1.037	> .20
8 - 6	.673	.612	1.100	> .20
9 - 10	.246	.696	.353	> .50
9 - 11	.793	.736	1.077	> .25

*One-tailed test

For an overall check on this, all classroom practice groups were combined and compared to classroom practice plus evening groups combined. The mean for classroom practice alone was 35.660 and for classroom practice plus evening viewing was 35.251, giving a difference of .409. The critical ratio for this difference was 1.023, which has a probability of about .30.

A second viewing in the evening plus classroom practice was significantly better than a second viewing alone (1). The difference in means was 3.194, which had a critical ratio of 3.834 and a probability less than .001.

Classroom practice plus evening viewing and parent help (groups 9, 10, and 11) gave results closer to what was expected. In two of the three comparisons, parent help groups did better, though for the eclectic groups (4 and 11) classroom practice alone was slightly better. Again none of the differences (9-2, 10-3, 4-11) was significant. The difference between means of combined groups, however, was significant. In this comparison, the classroom practice plus evening viewing groups were combined with classroom practice alone, since the evening viewing made no significant difference. The mean of classroom practice groups was 35.431 and that of classroom practice plus parent help groups was 36.118. The difference of .757 had a critical ratio of 2.151, significant at the .03 level.

Classroom practice plus evening viewing and parent help was far better than evening viewing and parent help alone (5). Pupils in the latter category had a mean of 33.001. The difference in means of 3.187 gave a critical ratio of 4.814 and a probability less than .001.

A major concern was the comparison of the three types of classroom practice. In both tests administered during the first semester -- the December mid-term and the January final -- the eclectic practice group did best, the structure practice group was second, and the dialogue practice group was third. On the January test, the difference between eclectic and dialogue was significant at less than the .05 level on a one-tailed test.

In the present analysis, which accounts only for second semester learning, the order of groups was the same for classroom practice alone and for classroom practice plus evening viewing as before. None of the differences (3-2, 4-2, 7-6, 8-6) was significant, however. When evening viewing groups were combined with classroom practice groups, which was logical since evening viewing had no significant effect, the difference between eclectic practice groups and dialogue practice groups was significant at the .076 level on a one-tailed test. The difference between means for these groups was .692 and the critical ratio was 1.439.

Among groups with classroom practice and parent help, however, the order of groups was reversed, with dialogue practice (9) best, structure practice (10) second, and eclectic practice (11) third. No difference was significant; however, the probability of the greatest difference (9 - 11) was greater than .25. The relative standings can therefore be attributed to chance. The possibility of an interaction between classroom practice and evening activities existed, however, and required further checking before any conclusions could be drawn.

A two-dimensional covariance analysis was run with classroom practice as the first effect and evening activity as the second effect. The January test result was used as the only control variable in this analysis since it was found in the one-dimensional analyses to account for practically all the variance attributable to the other control variables. The two-dimensional analysis is given in table 14.

Table 14

TWO-DIMENSIONAL COVARIANCE ANALYSIS
COMPARING CLASSROOM PRACTICE AND EVENING ACTIVITIES

Source of Variation	Residual			Frequency Ratio	Probability
	Degrees of Freedom	Sum of Squares	Mean Square		
First Effect	2	48.791	24.396	1.377	> .10
Second Effect	2	338.747	169.374	9.562	< .001
Interaction	4	137.661	34.415	1.943	> .10
Within Groups	1785	31,617.281	17.713		

N = 1795

This analysis shows that, with the three types of evening activity combined, no significant variation was produced by different types of classroom practice (first effect). With types of classroom practice combined, however, different evening activities did produce a significant variation (second effect). These results agree with comparisons in table 13.

More important, the interaction was not significant. Therefore, the inconsistency among classroom practices when combined with different evening activities, which was noted earlier, must be attributed to chance fluctuation.

The practices followed by groups one through five continued unchanged throughout the school year. A more valid comparison of these groups, then, would reflect the complete year's learning. That is, the second semester final should not be controlled for the January test. Group means for the whole year, controlled only for IQ, grade point average, and word meaning score, are shown in table 15. Results for the groups who had classroom practice the whole year and evening viewing the second semester are included since the evening viewing had no significant effect.

Table 15

GROUP MEANS REFLECTING LEARNING FOR THE WHOLE YEAR

	Group							
	1	2	3	4	5	6	7	8
Mean	28.117	35.178	35.807	35.804	33.141	35.028	36.058	35.980

Significance tests for differences between means are given in table 16.

Table 16

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS WHERE LEARNING FOR THE WHOLE YEAR IS REFLECTED

Groups Compared	Difference	σ_{D_M}	Critical Ratio	Probability
5 - 1	5.024	.817	6.149	< .001
2 - 5	2.037	.838	2.431	= .015
3 - 2	.629	.790	.796	> .20
4 - 2	.626	.771	.812	> .20
7 - 6	1.030	.619	1.664	= .047*
8 - 6	.952	.612	1.556	= .060*
7 - 8	.078	.602	.130	> .50

* One-tailed test

Results for the year are similar to those attained on both the first semester tests. The parent help group (5) performed significantly better than the group with a second viewing (1), and the lowest classroom practice group (2) was significantly better than the parent help group.

Dialogue practice groups (2 and 6) were poorest. The structure practice groups (3 and 7) and the eclectic practice groups (4 and 8) had almost identical results.

In all three tests the differences between structure practice and eclectic practice was slight, with the probabilities greater than .50, indicating chance fluctuation and acceptance of the null hypothesis.

The differences between dialogue practice groups and structure and eclectic practice groups were not significant for classroom practice alone (2, 3, 4), but they were significant where classroom practice was combined with evening viewing (6, 7, 8).

In view of the fact that evening viewing made practically no difference to groups with classroom practice, combining evening viewing groups with classroom practice groups in testing results for the whole year seemed logical and desirable. When this was done, the dialogue practice group had a mean of 35.086, the structure practice group a mean of 35.949, and the eclectic group a mean of 35.917. The difference between means of the structure and dialogue practice groups was .863 which had a critical ratio of 1.768, significant at the .038 level on a one-tailed test. The difference between eclectic and dialogue practice group means was .831, giving a critical ratio of 1.728 and a probability of .042 on a one-tailed test. The difference between structure practice and eclectic practice groups again showed chance fluctuation with a probability greater than .50.

A test of oral ability was also administered at the end of the second semester. Because of time limitations and the difficulty of evaluating large numbers of pupils, testing was restricted to the five "pure" groups, i.e., those who followed the same practice through the complete year. Forty pupils were randomly selected from each group, giving a total N of 200.

The test consisted of three parts: phonetic accuracy, structure, and fluency. Phonetic accuracy was defined as the pupil's ability to pronounce Spanish sounds properly, structure as his ability to place language elements in their proper syntatic order, and fluency as his ability to use the language to converse and to describe a simple scene. The maximum number of points on the test for each section were: 20 for phonetic accuracy, 4 for structure, and 32 for fluency.

Table 17 gives group means on each part of the test plus a total score which is a sum of the parts.

Table 17

GROUP MEANS ON SEPARATE PARTS
OF THE ORAL TEST

Group	Test Part			
	Phonetic Accuracy	Structure	Fluency	Total
1	17.60	1.48	4.18	23.26
2	18.33	2.30	8.98	29.61
3	18.85	2.56	8.33	29.74
4	18.98	2.55	9.60	31.13
5	19.22	2.41	6.29	27.93
Total	18.60	2.26	7.48	28.34

The results are generally consistent with the listening comprehension tests. On total score, the groups ranked exactly the same as they did on each of the other tests, and the results are approximately the same on structure and fluency. The only major change occurred on phonetic accuracy. Here groups one through four are in the same order, but group five (the parent help pupils) has moved to the front.

A high relationship between speaking ability and listening comprehension is suggested by table 17. Pearson product moment correlations between the second semester listening comprehension test and each part of the oral test are as follows: phonetic accuracy, .358; structure, .088; fluency, .680; and total score, .713. Though all the correlations are positive, even the .713 for total score leaves about 50 percent of the variance not accounted for by listening comprehension. To a large extent a different skill is being measured.

Critical ratios of differences between group means, with the .05, .01, and .001 levels of significance indicated, are given in table 18.

Table 18

CRITICAL RATIOS OF DIFFERENCES BETWEEN GROUP MEANS ON THE ORAL TEST, BY TEST PART

Groups	Test Part			
	Phonetic Accuracy	Structure	Fluency	Total
2 - 1	1.21	2.00*	3.17**	3.19**
3 - 1	2.09*	2.93**	3.53***	3.99***
4 - 1	2.28*	2.90**	3.76***	4.33***
3 - 2	1.33	0.62	0.39	0.13
4 - 2	1.62	0.60	0.33	0.81
5 - 1	2.95**	2.62**	2.08*	3.30***
4 - 5	0.77	0.39	2.24	1.83
3 - 5	1.23	0.41	1.67	1.12
2 - 5	2.82**	0.27	1.74	0.78
4 - 3	0.33	0.01	0.80	0.78

*P < .05
**P < .01
***P < .001

c. Causal Factors
Related to Double Viewing and Parent Help

As mentioned before, calendars were given to pupils who viewed the lesson again at night (with or without parental participation) and they were told to indicate the number of evening telecasts viewed and the number of minutes spent practicing with parents each week. The calendars were examined in an attempt to isolate the influence of evening activities on amount learned. Since only a limited number of the calendars were returned, the validity of broad conclusions based on them may be questioned. The returned calendars did offer some suggestions, however. The mean test score of those children whose parents returned calendars was over three points higher than that of the groups as a whole, for example. This indicates outstanding achievement by this part of the parent group, but it also poses many questions for further study.

Table 19 shows the average percent of evening programs watched by the evening viewing alone and evening viewing plus parent help groups the first semester.

Table 19

AVERAGE PERCENT OF EVENING PROGRAMS WATCHED THE FIRST SEMESTER BY EVENING VIEWING AND PARENT HELP GROUPS

Period	Evening Viewing	Parent Help
Prior to First Semester Mid-Term	62.07%	72.51%
Complete Semester	57.34%	67.32%

Table 19 reveals that the parent help group watched the evening telecast about ten percent more in each period than the evening viewing group, and this could account to some degree for the superior learning of those with parent help. Whether or not this is true, however, depends on the relationship between evening viewing and test score. The product moment correlation between these variables was .289 for parent help pupils and .154 for evening viewing pupils, so the percent of programs watched accounted in part, at least, for the difference in performance of the two groups.

Table 19 also shows that both groups watched the evening lessons considerably less in the period between tests than they did prior to the mid-term. This is not surprising since the Christmas season, a period when many regular activities are interrupted, occurred between the tests. The decrease in viewing could, in fact, account for the relative drop in performance (indicated by tables 8 and 9) by parent help pupils on the first semester final. A drop in performance of the evening viewing pupils is also suggested, and this did not occur. However, correlations in the preceding paragraph show that watching the evening lessons had a greater effect on parent help pupils than on the others, so results are generally consistent.

Just why the telecasts had more effect on parent help pupils is undetermined. Evidently the presence of parents during the viewing at home was motivating in some way. Motivation is also suggested by the larger percent of watching.

Test means, adjusted for evening viewing, turn out to be about the same for both groups. The superior first semester showing of the parent help group, then, appears to be largely the results of viewing the lessons with parents. This means, however, that practice with parents had little effect. Table 20 verifies this conclusion.

Table 20

MEAN HOME PRACTICE TIME AND CORRELATION BETWEEN PRACTICE TIME AND TEST SCORE THE FIRST SEMESTER FOR THE PARENT HELP GROUP

Period	Correlation	Average Minutes Practiced per Week
Prior to First Semester Mid-Term	.030	19.83
Complete Semester	-.023	30.46

A near-zero correlation was found between practice time and test score. This does not mean, however, that practice time had no indirect effect. As mentioned above, increased motivation is apparent in these results, and family practice undoubtedly contributed to it. Unfortunately, this must remain an assumption since there is no direct way to measure motivation.

Second semester figures give similar results, although the classroom practice plus evening viewing pupils continue to defy explanation. Table 21 shows average percent of programs watched and the correlation between programs watched and test score for groups with evening activities during the second semester.

Table 21

AVERAGE PER CENT OF EVENING PROGRAMS WATCHED
AND CORRELATION BETWEEN PROGRAMS WATCHED AND TEST SCORE
DURING THE SECOND SEMESTER

Group	Per cent of Programs Watched	Correlation
Evening Viewing	47.42%	.201
Parent Help	60.46%	.256
Evening Viewing + Classroom Practice	49.01%	.254
Parent Help + Classroom Practice	60.38%	.082

The groups without parent participation watched almost the same percent of programs and those with parent participation also watched the same percent. The correlation between programs watched and test score remained about the same for the parent help group as it was the first semester. For the evening viewing group it came up a little, and for the parent help plus classroom practice group it dropped considerably. The drop was expected, since classroom practice entered as a new factor. The high correlation for the evening viewing plus classroom practice group is surprising. Since these pupils watched about half of the programs, the correlation should mean a better performance than those with classroom practice alone; however, table 12 shows a slight decrease in learning.

Table 22 shows an increase in the relationship between practice time and test score for parent help pupils during the second semester.

Table 22

MEAN HOME PRACTICE TIME AND CORRELATION BETWEEN PRACTICE TIME AND TEST SCORE FOR PARENT HELP GROUPS DURING THE SECOND SEMESTER

Group	Correlation	Average Minutes Practiced per Week
No Classroom Practice	.220	7.59
Classroom Practice	.170	17.70

The correlation between practice time and performance was much higher than during the first semester. Perhaps parents and children had learned to work together more usefully; they were given additional information on practice procedures, and this could have helped. At any rate, the higher correlation is encouraging. If it should continue as a trend, the value of parent help would increase from year to year.

Parent help pupils without classroom practice registered a sharp drop in weekly amount of practice time compared to the first semester. Those with classroom practice showed a small drop but continued to practice over twice as much as the others, indicating that the classroom work encouraged more home activity.

d. Summary of Findings

The analyses reported in this section may be summarized as follows:

1. Two viewings of the television (i.e., repetition) are significantly superior to a single viewing.
2. Dialogue classroom practice, used in conjunction with the television series, is significantly less effective than eclectic classroom practice.

3. Eclectic classroom practice appears to be superior to structure classroom practice, although differences between means for these groups were not statistically significant.
4. Watching the television program again at night and having parent help is significantly superior to only watching again at night.
5. Each of the classroom practices is significantly superior to a second viewing and parent help at night but no classroom practice.
6. Combining classroom practice with a second viewing and parent help is significantly better than any other type of instructional pattern used.
7. Combining classroom practice with a second viewing at night (but no parent help) is not significantly different from classroom practice alone. In fact, groups with the evening viewing made scores slightly lower than those with classroom practice alone (no significant difference, however).
8. Pupils with parent help learn more from just watching the TV lesson again at night than do pupils without parent help, and they watch a larger percent of programs. This carries strong implications of a motivational effect connected with parent help.
9. The value of home practice with parents (exclusive of TV viewing) appears to increase with time.

IV. DISCUSSION OF FINDINGS AND OF FUTURE PLANS

The first year's research, to some extent at least, tested all of the hypotheses stated on page three, and in each case the hypothesis was verified.

Hypothesis one stated that instructional television reinforced by classroom activities would be more effective than television alone. On every evaluation during the year, the pupils with classroom practice learned significantly more than those without classroom practice.

Hypothesis two stated that home activities would increase the amount learned from television, and again every evaluation showed those with home activities had learned significantly more than those who merely viewed television once during school.

The conclusiveness of these results means that research related to the first two hypotheses may be considered complete. Large differences, with probabilities consistently less than .001, leave no doubt of the validity of findings.

Though hypotheses three and four were partially verified, a number of important questions related to them remain unanswered.

Hypothesis three, for example, stated: "Instructional television of a certain kind, reinforced by certain home and classroom activities, makes for more learning, ceteris paribus, than without these reinforcements or with either one alone." Classroom practice combined with evening viewing and parent help did prove significantly better during the second semester than anything else tried, and it supports the hypothesis. On the other hand, one would expect the combination of classroom practice and evening viewing also to support the hypothesis, but it did not. The reasons for this must be determined before any final conclusions regarding the most effective combination of classroom practice and home activities can be drawn. In addition, certain types of activities are yet to be tried.

This leads to hypothesis four, which states that certain kinds of home and classroom activities will be more effective than other kinds and combinations when used as reinforcement for instructional television. As far as the research has gone, hypothesis four holds. Eclectic classroom practice appears to be the most effective classroom activity; watching the evening telecast with parents and practicing with them is definitely more effective as a home activity than merely watching; and combining a classroom activity with evening viewing and parent help is more effective than any other combination or any single activity.

These results are not conclusive because of the number of possibilities to be tried during the remaining two years of research. Some of these are completely new and some are practices and combinations based on findings to date.

Regarding the latter, eclectic has been chosen for use in all classrooms and dialogue drill and structure drill eliminated as single methods. Parent help materials have been improved, largely along lines suggested by the first year's participants, and the method has been broadened so that any parent in any group may volunteer. Every pupil in the project will be encouraged to participate in some home activity in addition to classroom procedures.

New activities include the use of electronic aids such as tape recorders and record players. During the second year of instruction (sixth grade) pupils will learn to read and write the Spanish already covered thoroughly by the audio-lingual method and will engage in in-school activities not necessarily limited to the classroom. The in-school activities might include such things as a Spanish table in the lunchroom, Spanish games on the playground, and a Spanish alcove where items of cultural interest would be displayed.

The research design for the project will continue to be determined in the manner so far utilized. That is, the procedures used each year will be based upon previous results to a large extent, but always with efforts toward discovering more effective and refined activities.

This report covers approximately one-third of the time allotted to the project. Results of the remaining research will be made available as soon as possible after each school year.

REFERENCES

Andrade, Manuel, John L. Hayman, Jr., and James T. Johnson, Jr. "Measurement of Listening Comprehension via Television in Elementary School Spanish Instruction." Denver-Stanford Project on the Context of Instructional Television. Report number 3. Denver, Colorado: Title VII Office, October, 1961. (Mimeo.)

Brooks, Nelson. _Language and Language Learning_. New York: Harcourt, Brace and Company, 1960.

Carpenter, C. R. "Approaches to Promising Areas of Research in the Field of Instructional Television." In _New Teaching Aids for the American Classroom_. Stanford, California: Institute for Communication Research, 1960.

Hayman, John L., Jr. "A Comparison of Three Presentation Methods in Educational Television Instruction." Unpublished Ph.D. dissertation, Stanford University, 1961.

Hayman, John L., Jr., and James T. Johnson, Jr. "A New Opportunity for Parent Participation in Education." Denver-Stanford Project on the Context of Instructional Television. Report number 1. Denver, Colorado: Title VII Office, April 1961. (Mimeo.)

Runyon, Richard P. and Joseph H. Kanner. "Present Status of Signal Corps Television Research." _Audio-Visual Communication Review_. 4 (1956), 83-91.

Schramm, Wilbur. "What We Know about Learning from Instructional Television." Stanford, California: Institute for Communication Research, 1961. (Mimeo.)

THE AUTHORS

John L. Hayman, Jr., is director of the Denver-Stanford project for Stanford University and is specifically responsible for research methodology and analysis in the project.

James T. Johnson, Jr., is director of the Denver-Stanford project for the Denver Public Schools and serves as overall coordinator of the project.

ED035269

READING AND WRITING RESULTS IN THE SECOND YEAR OF RESEARCH—1961-62

• JOHN L. HAYMAN, JR.
• JAMES T. JOHNSON, JR.

EM007066

Denver Public Schools • Stanford University

RESEARCH ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

ED035269

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

READING AND WRITING RESULTS IN THE SECOND YEAR OF RESEARCH--1961-62

by

John L. Hayman, Jr.
James T. Johnson, Jr.

DENVER-STANFORD PROJECT ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

****

School District Number One
City and County of Denver
Denver, Colorado

Institute for Communication Research
Stanford University
Stanford, California

Report Number 7
May 1963

SUMMARY

The Problem

Fifth grade pupils in the Denver-Stanford project study Spanish entirely by the audio-lingual method. At sixth grade, however, they learn to read and write the vocabulary and functional grammar used orally the first year.

Project personnel have sought to answer two major questions relative to development of the reading and writing skills. First, when is the best time to begin reading and writing instruction? Foreign language experts disagree as to whether this instruction should be introduced at the beginning of sixth grade or at the beginning of the second semester. Second, what is the best method for teaching reading and writing? Could it be handled best through the traditional teacher-directed approach or through the newer medium of automated instruction?

Results

Sixth grade pupils were randomly divided into two groups. The first group started reading and writing at the beginning of the first semester, and the second group started at the beginning of the second semester. These groups were, in turn, randomly divided, and half of each studied reading and writing through automated instruction, while the other half studied through the traditional teacher-directed approach.

Concerning the best time to begin reading and writing instruction, pupils who had this instruction the first semester performed better on tests of understanding and speaking skills at the end of the semester than pupils who did not have reading and writing. And, at the end of the year, pupils with reading and writing the first semester again performed significantly better on understanding and speaking tests than pupils who started reading and writing the second semester. Since the objection to starting reading and writing at the beginning of the year was based on the fear that it would adversely affect the speaking skills, and obtained results showed this definitely did not occur, the earlier beginning of reading and writing instruction appears desirable.

The comparison of teaching methods was based on relative performances on a test of reading and writing skills. At the end of the first semester, pupils with the teacher-directed approach performed significantly better than those with automated instruction. At the end of the second semester, however, there was no significant difference between pupils who studied by the different methods. This was true for pupils who started reading and writing at the beginning of the year and for those who started the second semester. (Second semester results for the former group were adjusted by the covariance technique for first semester performance, so that only the amount learned during the second semester was reflected.) These results suggest an important timing effect in the use of automated instruction. Any automated program must necessarily assume a certain prior knowledge, and, for the program to be most effective, users must possess this knowledge. A teacher can make individual adjustments if a pupil is not fully prepared for the instruction at hand, but an automated program cannot. Within this limitation, automated instruction can be highly effective.

Another finding relative to the automated instruction was that the training and experience of the classroom teacher who taught the audio-lingual skills had a significant effect on the amount pupils learned from automated instruction in reading and writing. In automated instruction, pupils are on their own to a great extent, and their desire to learn largely determines what they do. Thus the classroom teacher, who has a great effect on this desire, helps determine the performance of pupils in automated instruction.

Finally, two versions of the automated materials were used. Each was a linear program and used the same frames, but the sequence of frames was different. One sequence used what were apparently minimum learning steps, while the other used larger learning steps. The sequence with minimum learning steps produced significantly fewer response errors, took significantly less time to complete, and produced significantly superior results on the reading and writing test.

READING AND WRITING RESULTS IN THE SECOND YEAR OF RESEARCH--1961-62

by

John L. Hayman, Jr.
James T. Johnson, Jr.

The Denver Public Schools and Stanford University's Institute for Communication Research are currently engaged in a joint research project on the context of instructional television. The purpose of the project is to learn how instructional television can best fit into the total teaching situation. A substantial amount of research has established that television is a very effective teaching medium. Ways of combining it with other educational activities must now be considered, and the Denver-Stanford project is a beginning effort in this direction. Kenneth E. Oberholtzer is principal investigator for the Denver Public Schools and Wilbur Schramm is principal investigator for Stanford University. This is one of a number of project progress reports.

The Denver-Stanford project is concerned with teaching beginning Spanish to some 12,000 fifth and sixth grade pupils. All pupils view televised Spanish lessons and engage in teacher-directed classroom activities designed to complement the TV lessons.

In keeping with current methodology in elementary school foreign language instruction, audio-lingual practice is used exclusively during the first year of study (fifth grade). During the second year (sixth grade) the development of audio-lingual skills is continued, but reading and writing are also introduced. The vocabulary and functional grammar used in reading and writing are limited to material which is taught audio-lingually in the fifth grade and with which pupils have oral facility.

Project personnel have sought to answer two major questions related to reading and writing instruction. First, when is the best time to introduce reading and writing -- at the beginning of sixth grade after two semesters of exclusive audio-lingual work or at the beginning of the second semester of sixth grade when pupils have worked three semesters exclusively on the development of understanding and speaking skills? This question is cogent because the speaking skills, especially, can be adversely affected if pupils see the written word too soon. A premature introduction of reading and writing can lead to anglicization.

The second question concerns the best method of teaching reading and writing. Traditionally, reading and writing instruction has been handled by the classroom teacher in a face-to-face, group-learning situation. The traditional method has proved effective over many years, of course. The recent development of automated instruction and the dramatic experimental results it has produced, however, raise the question as to whether the reading and writing instruction given in the Denver-Stanford project could not be handled as effectively through the automated approach as with teacher-direction.

Both programed and conventional reading-and-writing texts were developed during the 1960-61 school year so that a comparison of the two instruction methods could be made. The development process is described in a previous report (1).

This report deals with reading and writing results during the second year of research in the Denver-Stanford project -- the 1961-62 school year.

Research Design for Reading and Writing

At the beginning of the year, sixth grade pupils were divided into two groups, each consisting of approximately 90 classes with some 3,000 pupils. One group started reading and writing at the beginning of the school year, while the other group started at the beginning of the second semester. This division was to allow evaluation of different starting times for reading and writing instruction.

Each of the two major groups was in turn subdivided into two parts, with one subgroup learning to read and write through the traditional teacher-directed method and the other learning through automated instruction. Programed textbooks were used by all pupils receiving automated instruction during the first semester and by all but four classes during the second semester. These four classes used a commercially-manufactured teaching machine.

Two versions of Automated Spanish, the panel text, were used. The first -- the I-II series -- consisted of the original arrangement of frames which was developed through work with individual pupils; the second -- the A-B series -- was a different arrangement of the original frames based on suggestions by staff members of the Center for Programed Instruction. The A-B series was less sequential than the I-II series, and it therefore involved larger learning steps.

This research design allowed answers to these questions:

1. Does the introduction of reading and writing, after only one year of audio-lingual instruction, have any effect on the listening and speaking skills of pupils?
2. How does automated instruction of reading and writing compare with instruction by the more traditional teacher-directed method.
3. In automated instruction, do pupils learn more from a minimum step sequence of frames, or do they learn more from a larger step but supposedly more interesting sequence?

Measuring Instruments

Two reading and writing tests, final exams for each semester, were developed for this phase of the study. As given, the 66-item first semester final had an average biserial correlation (the index of discrimination) of .713 and a split-half reliability of .938. The second semester final, with 50 items, had an average biserial correlation of .593 and a split-half reliability of .957. The Pearson product moment correlation with listening comprehension tests was .634 for the first semester final and .684 for the second semester final.

As mentioned in an earlier report, a training and experience questionnaire was completed by all classroom teachers during the second semester. On this questionnaire, teachers indicated their foreign language courses in high school and college, their years of experience in teaching foreign language, any foreign language workshops they had attended, and any experience either in a bilingual home or in a foreign country with another language.

In developing a scale from questionnaire results, fifth grade teachers were used, and the first semester listening comprehension results from their classes constituted the criterion. These teachers were randomly divided, and half of them were used in the scale development. The scale was then cross-validated with the other half of the teachers.

Formal training was arranged in five categories as follows:

1 = none
2 = a foreign language other than Spanish in high school but no further training
3 = Spanish in high school but no further training
4 = one year of foreign language in college
5 = two or more years of foreign language in college

Experience was arranged in the following three categories:

1 = none
2 = taught Spanish one or two years or taught French one or more years
3 = taught Spanish three or more years

And workshop attendance was arranged in two categories:

0 = have never attended a foreign language workshop
1 = have attended one or more foreign language workshops

In deriving a scale score for each teacher, formal training was weighted three, experience was weighted one, and workshop attendance was weighted one. The scores, therefore, had a possible range of 5 to 20. Bilingual experience did not seem to have any effect on pupil performance and was not used in the scale.

Teachers were assigned scale scores and then arranged in three groups, which were classified low (5 to 8 scale points), middle (9 to 14 scale points), and high (15 to 20 scale points) on training and experience. The efficacy of the scale is demonstrated in table 1, which gives mean scores of fifth and sixth grade pupils on second semester tests.

Table 1

PERFORMANCE OF PUPILS
WHEN CLASSIFIED ACCORDING TO
THE TRAINING AND EXPERIENCE OF THEIR TEACHERS

		Test		
		Fifth Grade Listening Comprehension	Sixth Grade Listening Comprehension	Sixth Grade Reading and Writing
Teacher Training and Experience	High	36.750	32.671	18.049
	Middle	35.231	32.049	15.579
	Low	33.069	31.154	12.636

All differences between vertical categories in table 1 are significant beyond the .05 level.

Results

The First Semester

a. Comparison of Learning Methods

First semester results were analyzed by a one-dimensional covariance analysis in which the 66-item final was the dependent variable, and grade point average, IQ, reading and writing pre-test, and listening comprehension pre-test were control variables. The six groups compared, with number of subjects from each group in the analysis are shown in table 2.

It will be recalled that the I-II series was the sequence of frames originally developed, and the A-B series was the changed sequence based on suggestions by the Center for Programed Instruction. The changed sequence was suggested as a means of testing the point of view that a larger learning step would prove more interesting and would therefore be a more effective teaching medium, at least for some pupils.

Table 2

EXPERIMENTAL GROUPS AND NUMBERS OF SUBJECTS IN ANALYSIS OF READING AND WRITING TEST

Experimental Group		Number of Subjects in Analysis
Automated, Book 1	(A_1)	93
Automated, Book A	(A_A)	169
Teacher-Directed	(TD)	193
Automated, Book 1 + Electronic Aids	(A_1+EA)	277
Automated, Book A + Electronic Aids	(A_A+EA)	171
Teacher-Directed + Electronic Aids	(TD+EA)	222
	Total	1125

Adjusted test means plus standard deviations and variances are given in table 3.

Table 3

ADJUSTED TEST MEANS, STANDARD DEVIATIONS, AND VARIANCES FOR EACH EXPERIMENTAL GROUP

	Group					
	Automated$_1$	Automated$_A$	Teacher-Directed	Automated$_1$+ Electronics	Automated$_A$+ Electronics	Teacher-Directed+ Electronics
Reading and Writing Test	17.001	15.725	17.665	18.020	16.132	22.167
σ	14.077	12.048	14.097	12.492	12.328	17.077
σ^2	198.167	145.167	198.745	156.071	151.985	291.646

Results of the covariance analysis of these data are given in table 4. The probability of less than .001 indicates that real differences occurred among the groups, and group comparisons are in order. The test means in table 3 show a very consistent pattern. In each case the I-II series is superior to the A-B series; teacher-directed reading and writing is superior to the programed texts; and those with electronic aids did better than similar groups without electronic aids. Tests of the significance of differences between groups are given in table 5.

Table 4

COVARIANCE ANALYSIS OF DIFFERENCES BETWEEN EXPERIMENTAL GROUPS ON THE READING AND WRITING TEST

Source of Variation	Residual		
	DF	Sums of Squares	Mean Square
Total	1120	82,139.767	
Within Groups	1115	76,605.807	68.705
Between Groups	5	5,533.960	1,106.792

$$F = \frac{1{,}106.792}{68.705} = 16.109$$

$n_1 = 5 \qquad n_2 = 1115$

$P < .001$

Table 5

SIGNIFICANCE OF GROUP DIFFERENCES ON THE READING AND WRITING TEST

Groups	Difference	σ_{D_M}	CR	P
Automated$_1$ - Automated$_A$	1.834	0.949	1.933	<.06
Teacher-directed - Automated	3.189	0.918	3.474	<.001
Electronic Aids - No Electronic Aids	2.101	0.869	2.418	<.02

These results leave no doubt that pupils with the teacher-directed method learned more during the first semester. Several other questions need consideration before any final conclusions can be drawn, however. Would these results be consistent for teachers with limited training and experience as well as for well-trained teachers? What would a comparison between pupils of different IQ levels show? What about the pupils who worked rapidly through the automated books as compared to those who worked slowly?

Rather surprising was the finding that electronic aids seemed to help pupils learn Spanish reading and writing. The electronic aids were expected to improve listening and speaking skills but not reading and writing.

Finally, the original sequence of frames apparently worked better than the larger-step sequence proposed by CPI.

Means for the combined groups were: I-II series -- 17.764; A-B series -- 15.930; all automated -- 16.885; teacher-directed -- 20.074; electronic aids -- 18.893; no electronic aids -- 16.792.

b. Characteristics of the Automated Programs

Extensive records were kept during the first semester so that such characteristics of the programs as error rate and time rate could be examined. An analysis of these data indicates that the rearrangement of frames did indeed make a difference in pupil performance on the programs.

Regarding error rate, the original sequence (the I-II series) was written so that pupils made as few errors as possible. That is, a minimum error rate was sought. The altered sequence proposed by CPI (the A-B series) involved larger learning steps between some frames, and the error rate was therefore expected to increase, though there was no assurance that it would.

The A-B series, in fact, did have a higher error rate. Pupils with this book made an average of 47.648 errors in completing their first 672 frames for an error rate of .071. Pupils using the I-II series on the other hand, averaged 38.353 errors on the same number of frames for a rate of .057. The difference of 9.295, with σ_D = 3.520, gave a critical ratio of 2.641, which is significant beyond the .01 level. In other words, the A-B series had a significantly higher error rate than the I-II series. This is clearly seen in table 6; more than twice as many of the I-II's made four errors or less than did the A-B's, while more than twice as many A-B's than I-II's made more than 150 errors.

Table 6

ERROR DISTRIBUTION AMONG
BOOK I AND BOOK A PUPILS

Number of Errors	Pupils Making This Many Errors Book I	Book A
0 - 4	16.3%	7.7%
5 - 9	8.7	10.4
10 - 49	49.6	51.5
50 - 99	17.8	17.4
100 - 149	5.5	7.5
More than 150	2.0	5.6
	100.0%	100.0%
(N)	(343)	(483)

The range in error rates was surprisingly large. A few pupils made no errors (they did not mark any, at least), while almost two percent made more than 200. One child marked himself wrong 351 times as he went through the first 672 frames, giving him an error rate of .522.

Error rate correlated -.285 with results of the first semester final test -- that is, pupils with lower error rates tended to make higher grades on the test. Though one must be cautious in ascribing cause and effect, this negative relationship fits neatly into the pattern so far observed. The A-B series produced more errors than the I-II series; pupils with fewer errors tend to make higher scores; and, as would be predicted, the I-II's learned significantly more than the A-B's. This suggests that, as Skinner and his followers have held, a minimum error rate is desirable.

A wide range was also found in the amount of time required to complete the first book (672 frames). One pupil finished on November 11, only seven weeks after starting the program; while two pupils did not finish until April 11, after 26 weeks. Since all pupils used the automated texts only in class, for a standard work period of 30 minutes a week, these completion times represent a range from 3½ to 13 hours required to work through the first book.

With this wide range in working time, pupils will obviously require widely different amounts of materials in a school year. For all pupils, the average number of weeks required for the first book was 14.03, so three books would suffice for a 36 week year on the average. However, 24.5 percent of the pupils finished in 11 weeks or less and would require more than three books, and 1.5 percent finished in less than nine weeks and could use a fifth book.

Book I and Book A required different amounts of time for completion. For Book I, 13.688 weeks were required; whereas, 14.338 weeks were required for Book A. The difference, .650 weeks or about 19.5 minutes working time, gave a critical ratio of 1.947, which is significant at about the .05 level.

Length of time required to complete the first book, like error rate, showed a significant negative correlation (-.255) with test score. Table 7 shows this relationship clearly.

Table 7

MEAN SCORES OF PUPILS
WHO COMPLETED THE FIRST BOOK
AT DIFFERENT TIMES

Weeks Spent On First Book	Percent of Pupils In This Category		Mean Score on First Semester Test	
	Book I	Book A	Book I	Book A
11 or less	28.6%	21.7%	20.292	18.743
12 - 13	22.6	27.3	15.158	14.682
14 - 16	33.3	27.3	14.018	14.909
17 or more	15.5	23.6	7.615	9.342
	100.0%	100.0%		

The correlation between time and score may be somewhat spurious, of course, because some pupils did not complete all material covered in the test. Less than 25 percent fell in this category, however, and the top three categories in table 7 still show a strong relationship. Even if failure to complete part of the work is a contributing factor, the time requirement is still important; if two programs can teach the same amount, the one requiring less time would seem desirable.

Both the correlations of error rate with test score and time with test score might be caused by a third variable -- ability. That is, more able pupils might take less time and make fewer errors, and the relationships found would therefore be merely reflections of the correlation between ability and test score. In this case, time and error rate would be measures of the same thing and would correlate highly with each other. But since the correlation between error rate and time required was only .167, they are not just separate indicators of ability. The multiple correlation of error rate and time required with test score is -.355, which represents a significant increase over either variable used singly in amount of variance accounted for. (Using the F ratio test on page 279 of McNemar's Psychological Statistics, the increase in correlation was significant beyond the .001 level.)

These results, we believe, make a strong case for the linear, minimum learning step sequence as used in Book I.

The Second Semester

During the second semester, pupils who began reading and writing at the start of the year continued with their respective methods, and the remainder of the sixth graders began reading and writing. As previously mentioned, the new pupils were divided into automated and teacher-directed groups and a new comparison was possible since four classes from the automated group actually worked with teaching machines.

In addition, the completion and administration of the teacher training and experience scale made a much more comprehensive analysis possible at the end of the year. Second semester results, divided into three teacher training and experience groups, are shown in tables 8 and 9. The reader should remember that these tables show results of different tests. Pupils who began reading and writing the second semester took the 66-item first semester final at the end of the year, while those who began at the start of the year took the 50-item second semester final. Also, these tables reflect only second semester learning; results in table 8 are adjusted for the first semester final.

Both tables show the marked effect of teacher training and experience on learning. Differences on this variable produced an F ratio of 7.650 for table 8 and 10.613 for table 9, and with n_1 of 2 and n_2 of 696 and 952 respectively, both F ratios are significant far beyond the .001 level. Even more interesting is the fact that this effect holds among the automated pupils, except for those who actually used a teaching machine.

Both analyses also produced interactions significant beyond the .05 level. The source of the interaction is the horizontal dimension in table 8. That is, results are consistent vertically but inconsistent horizontally. Among pupils with teachers high on the training and experience scale, the teacher-directed method was best; but for pupils with teachers low on the scale automated instruction was best. This outcome seems entirely reasonable, and the temptation is strong to accept it as a general principle. Table 9 precludes such acceptance, however.

Table 8

ADJUSTED MEAN SCORES OF PUPILS
WHO STARTED READING AND WRITING THE FIRST SEMESTER --
ARRANGED BY LEARNING METHOD AND BY TEACHER TRAINING AND EXPERIENCE

		Learning Method		
		Automated Text	Teacher-Directed	Total
Teacher Training and Experience	High	15.174	16.598	15.753
	Middle	14.882	15.146	14.997
	Low	14.758	12.609	13.672
	Total	15.081	15.741	15.361

Table 9

ADJUSTED MEAN SCORES OF PUPILS
WHO STARTED READING AND WRITING THE SECOND SEMESTER --
ARRANGED BY LEARNING METHOD AND BY TEACHER TRAINING AND EXPERIENCE

		Learning Method			
		Automated		Teacher-Directed	Total
		Text	Machine		
Teacher Training and Experience	High	18.348	16.681	16.526	17.619
	Middle	15.227	19.288	13.994	14.993
	Low	12.763	20.036	17.112	14.812
	Total	16.414	18.439	15.447	16.165

In table 9, both dimensions contribute to the interaction. Teacher training and experience has no clear-cut effect on teacher-directed pupils, and it reverses the order for teaching machine pupils -- those with the best trained teachers did poorest and vice-versa. In addition, the comparison of automated text and teacher-directed pupils shows a reversed effect from table 8. Here the automated pupils did best when teachers were high on the training and experience scale, while teacher-directed pupils did best when teachers were low on the scale. So the suggestion in table 8 that automated instruction will show up best when compared to the work of less-well-trained teachers fails to hold up.

The teaching machine result in table 9 was very surprising to project personnel. The particular model used broke down consistently, and the children appeared to be repairing their machines at least a quarter of the time. In spite of this, children with the machines did significantly better ($P < .001$) overall than the others, and the breakdown by training and experience shows them better in every comparison except one (versus the automated text in the high category). This could be the result of "gimmick value," with the novelty of the situation stimulating learning for the relatively short period of the study. Or it could be that control of the responses made the difference, that it is important to prevent the possibility of seeing correct answers before responding overtly.

Finally, in both cases represented by tables 8 and 9, no significant difference occurred between pupils who learned from the automated texts and those who were teacher-directed. This is in sharp contrast to the first semester when teacher-directed pupils did far better than the others, and it suggests that timing was a very important factor in determining the relative merits of the two learning methods. In fact, timing could help account for the differing effects of teacher training, a point which will be discussed later.

First, however, consider the influence of IQ on the separate learning methods. Tables 10 and 11 show the results for both second semester and beginning pupils when arranged in high, middle, and low IQ groups.

Table 10

ADJUSTED MEAN SCORES OF PUPILS WHO STARTED READING AND WRITING THE FIRST SEMESTER ARRANGED BY LEARNING METHOD AND IQ

		Learning Method		
		Automated Text	Teacher-Directed	Total
IQ	High	14.271	16.005	14.926
	Middle	14.558	13.662	14.198
	Low	14.314	15.487	14.687
	Total	14.413	14.706	14.521

In both tables significant differences occurred among the different IQ groups, and in both the comparison of learning methods produced the same results as tables 8 and 9. (Overall means here are slightly different from those in tables 8 and 9 because random withdrawals from groups to meet the proportional cell frequencies assumption resulted in slightly different samples.) Table 10 produced an interaction significant at the .01 level, while there was no significant interaction for table 11.

Table 11

ADJUSTED MEAN SCORES OF PUPILS WHO STARTED READING AND WRITING THE SECOND SEMESTER ARRANGED BY LEARNING METHOD AND IQ

		Learning Method			
		Automated		Teacher-Directed	Total
		Text	Machine		
IQ	High	20.786	21.901	19.695	20.433
	Middle	15.708	18.391	16.128	16.024
	Low	15.360	15.860	13.848	14.799
	Total	16.506	18.949	16.088	16.471

The interaction in table 10 suggests no clear-cut conclusion. Among automated pupils, the middle IQ group did best, while this group did poorest among the teacher-directed pupils. The inconsistency can be seen in another way: teacher-directed pupils did best in the high and low IQ groups, while automated pupils did best in the middle group. Whatever these inconsistencies may mean, they do not support the finding from the English 2600 study that high IQ pupils learn more from automated instruction while low IQ pupils learn more when teacher-directed (2). Rather, IQ seems to have no overall effect except that higher IQ pupils generally do better with any form of instruction than lower IQ pupils.

During the second semester, in contrast to the first, electronic aids had no effect on reading and writing test results.

Discussion of Results

Our results suggest a number of conclusions relative to the use of automated instruction and to the teaching of reading and writing in Spanish.

First, we have definite indications that a linear, minimum learning step sequence is desirable for the type of learning involved here. When compared with a larger learning step sequence, the minimum sequence proved to produce fewer response errors, to take less time, and to result in more learning.

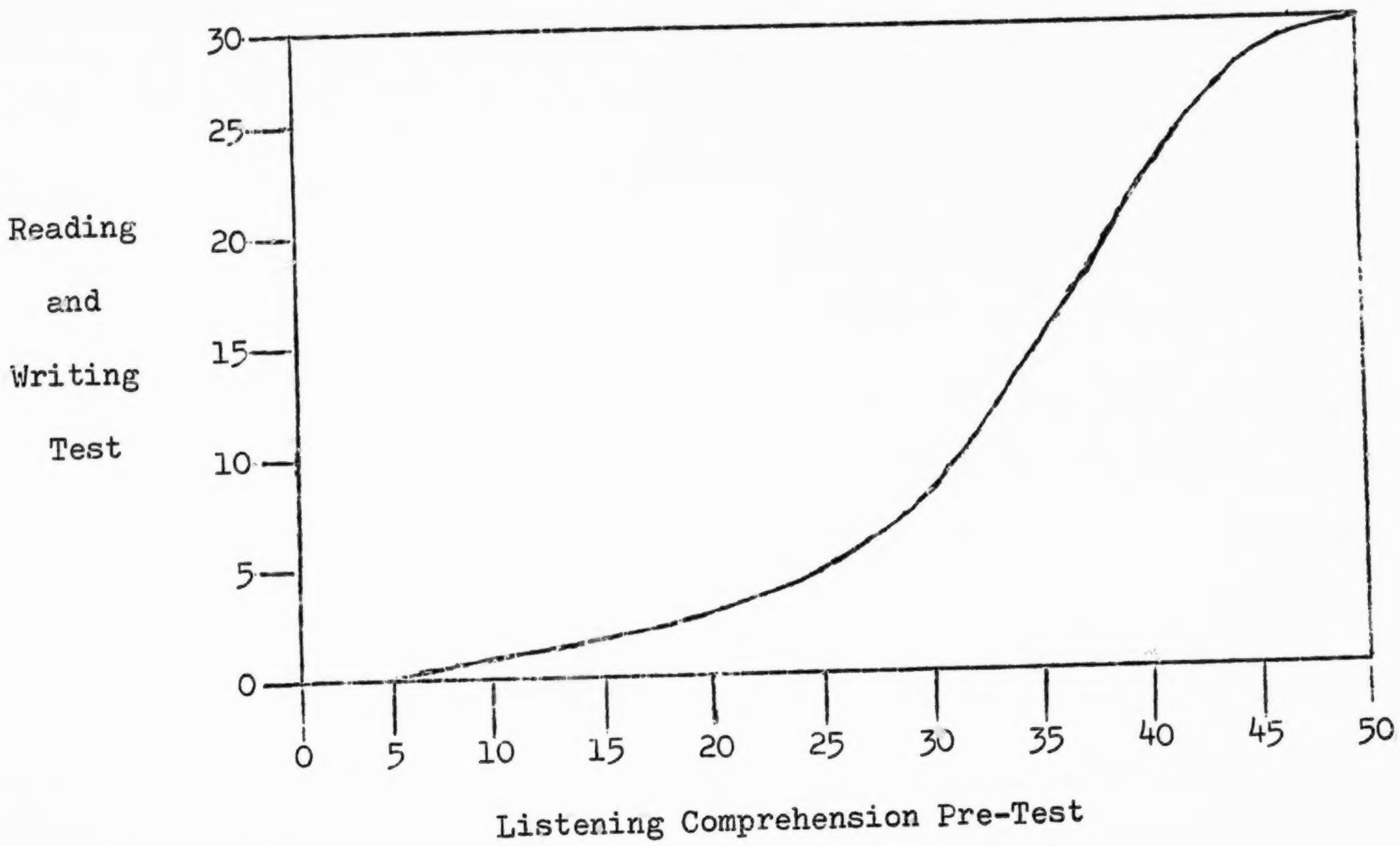
Another conclusion is that, with proper conditions and at least with types of learning which require a good deal of routine, repetitive work, automated instruction can be as effective as the more traditional teacher-directed method. In fact, the teaching machine results suggest that automated instruction can be superior, though conditions for this superiority are uncertain.

Timing appears important in determining the effectiveness of automated instruction. Our automated program assumed a certain prior knowledge at the audio-lingual level, and evidently many pupils did not possess this knowledge at the beginning of the first semester of sixth grade. Automated instruction proved less effective than teacher-directed instruction the first semester, but in two separate trials the two methods were of equal effectiveness the second. The timing or prior knowledge effect becomes even more apparent when reading and writing results are arranged according to listening comprehension pre-test scores. This arrangement is demonstrated both in table 12 and in the graph which follows.

Table 12

FIRST SEMESTER READING AND WRITING
RESULTS OF AUTOMATED INSTRUCTION PUPILS
ARRANGED BY LISTENING COMPREHENSION PRE-TEST SCORES

Listening Comprehension Pre-Test Score	Percent of Pupils	Mean Score on Reading and Writing
25 or less	17.7%	4.061
26 - 30	18.5	6.277
31 - 35	22.2	11.871
36 - 40	23.5	19.421
41 - 45	15.1	26.587
46 - 50	3.0	29.560
	100.0%	



The graph shows that pupils with a score of 25 or less on the listening comprehension pre-test do not do well in reading and writing. Somewhere between 25 and 30, however, the curve takes a sharp turn, and a small improvement in listening comprehension then makes a big difference in reading and writing. The slope decreases again at about the 45 level.

The critical point, then, seems to be about 28 to 30. If pupils do not have sufficient knowledge at the audio-lingual level to score this high on the listening comprehension pre-test, they will profit little from our automated instruction. Above this point, however, automated instruction is very effective.

Since almost any automated program will of necessity assume some knowledge and pupils will attain this knowledge at different times, the timing effect could well have broad implications. If a single program is to be used for a subject area, it should probably be introduced to different levels of students at different times. If automated instruction is to begin for all students at the same time, on the other hand, different programs should be used for different levels of students. In either case, the result would be a further separation among students in amount of knowledge possessed.

The difference in work rate, discussed in the section on first semester results, adds to this problem (if it is a problem). We found that some pupils would still be in the second book after one year's work, while others would have finished six books. This is a ratio of more than three to one. Applied to several subject areas and allowed to continue unabated, this difference in rate of progress would certainly disrupt our present educational system. It would no longer be possible to group pupils as first graders, second graders, etc., and move them along from year to year at the same general pace. In fact, little grouping would be possible, and teachers, instead of working in group learning situations, would work with pupils individually.

To say we are not presently ready for this is an understatement. Yet, we have proved that automated instruction is effective, and we should use it. The question is how? Obviously some modification of a strict automated approach is needed.

In another comparison, we found that classroom teachers had an effect on the amount pupils learned from automated instruction. In this situation, where students progress at their own rates, they are almost entirely on their own, and motivation to learn must be of great importance. Motivation is a complex affair with many contributing factors, of course, but there can be no doubt that the classroom teacher's contribution is of major proportion. Among pupils who started reading and writing the second semester, those with the best-trained teachers learned more from the automated than the teacher-directed method, while those with less-well-trained teachers did not do as well with automated. Evidently the better-qualified teachers inspired their pupils to learn more, and this inspiration showed up clearly when automated instruction put the pupils on their own.

Finally, the fact that many pupils were not ready for the automated materials at the beginning of the year does not mean that they were not ready to study reading and writing. Rather, two separate findings indicate they were ready. First, as we have reported previously, the reading and writing instruction had no adverse effect on either listening comprehension or speaking scores. And, second, the teacher-directed group did quite well the first semester; their mean reading and writing score was about 33 percent higher than that of the automated pupils. So the classroom teachers, by adjusting instruction to the needs of the children, were able to do quite well at the beginning of the year.

Overall, then, we have these results: automated instruction is effective when pupils are ready for it, teachers are also effective and can adjust the instruction to fit individual and group needs, and teachers affect the amount learned from the automated materials. This combination of findings suggests a possible solution to the problem posed above -- namely, that teacher-directed and automated instruction should be used together. By assigning specific portions of a program either as class or homework while using the group learning situation regularly, a teacher should be able to increase the amount children learn in a given time and still move a heterogeneous group at roughly the same rate. Since we are investigating various combinations of automated and teacher-directed instruction this year, we should know more about this possibility in the near future.

REFERENCES

(1) Barcus, Delbert, John L. Hayman, Jr., and James T. Johnson, Jr. "Development of Programed Learning Materials for Use with Televised Spanish Instruction." Denver-Stanford Project on the Context of Instructional Television. Report number 4. Denver, Colorado: Title VII Office, November, 1961. (Mimeo.)

(2) Reed, Jerry E. and John L. Hayman, Jr. "An Experiment Involving Use of English 2600, An Automated Instruction Text." The Journal of Educational Research. 55:476-84 (June-July, 1962).

THE AUTHORS

John L. Hayman, Jr., is director of the Denver-Stanford project for Stanford University and is specifically responsible for research methodology and analysis in the project.

James T. Johnson, Jr., is director of the Denver-Stanford project for the Denver Public Schools and is responsible for overall coordination of the project.

ED035269

THIRD YEAR RESULTS IN THE DENVER-STANFORD PROJECT

•JOHN L. HAYMAN, JR.
•JAMES T. JOHNSON, JR.

EM007066

Denver Public Schools • Stanford University

RESEARCH ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

ED035269

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THIRD YEAR RESULTS IN THE DENVER-STANFORD PROJECT

by

John L. Hayman, Jr.
James T. Johnson, Jr.

DENVER-STANFORD PROJECT
ON THE CONTEXT OF INSTRUCTIONAL TELEVISION

****

School District Number One
City and County of Denver
Denver, Colorado

Institute for Communication Research
Stanford University
Stanford, California

Report Number 10
March 1964

SUMMARY

The Problem

During 1962-63, the last year of active research in the Denver-Stanford Project, work was continued at both the fifth and sixth grades. As in previous years, fifth grade instruction was concerned entirely with the development of listening and speaking skills--that is, it was entirely audio-lingual in nature. Audio-lingual instruction was continued at sixth grade, and, in addition, development of reading and writing skills was begun.

The research at fifth grade was designed to determine, first, the value of certain electronic aids and, second, the most useful ways to incorporate these aids into the instructional process. Some children used no electronic aids. Some listened and responded to special lessons on record players and tape recorders but did not record their own voices. Still others both listened and recorded their own voices.

At sixth grade, the research was directed mainly toward evaluating methods of teaching reading and writing. Reading and writing was taught entirely by the traditional teacher-directed method, entirely by the automated method (programed instruction) and by various combinations of these two methods. Other activities included extended reading of simple Spanish materials and use of a "Spanish Corner," in which was placed reading materials, electronic aids, and cultural artifacts.

Finally, parent participation in the instructional process was continued at both fifth and sixth grades.

Results

Results at fifth grade suggested that electronic aids, especially those with feedback (that is, where the child records and listens to his own voice), are a desirable addition to the classroom Spanish program. An exception to this occurs, for reasons so far unexplained, among children from homes where Spanish is spoken natively. These children are very few in number, however, so that the electronic aids appear potentially helpful to the great majority of fifth grade pupils.

The sixth grade results confirmed a previously stated hypothesis that a combination of automated and teacher-directed reading and writing instruction would be superior to either method alone. The combination produced significantly superior results on both listening comprehension and reading and writing tests. Speaking test results showed the combination slightly less favorable, apparently because of less face-to-face contact between pupil and teacher when automated instruction is used. It appears, however, that this difficiency can be overcome through use of electronic aids outside of the regular Spanish instruction period.

Special interest scales were used at sixth grade, and they showed that both teacher and pupil interest affect learning. Pupil interest related, as a dependent variable, to two factors. The stronger was pupil interest prior to the beginning of the school year, emphasizing the importance of the pupil's experience at fifth grade. The second factor was experimental treatment; pupil interest was higher under the conditions which provided some activity in addition to those of the regularly scheduled classroom period. Homework was one such additional activity. Teacher interest was found to interact with teacher experience and preparation in a most revealing manner. Where teacher interest was high, pupil performance was directly related to teacher preparation and experience. But, where teacher interest was low, pupil performance was inversely related to teacher preparation and experience.

Class size was considered and smaller classes proved advantageous only in those situations involving direct teacher instruction.

Overall, the third year results reemphasized the importance of the classroom teacher, even when the basic instruction is handled through television as in this project. A well trained and highly motivated classroom teacher is undoubtedly the most effective single "learning aid" that a school system can combine with instructional television.

Finally, parent participation again proved a valuable addition to the instructional package at both fifth and sixth grades. Pupils whose parents worked with them learned significantly more Spanish than those whose parents did not participate.

TABLE OF CONTENTS

PART PAGE

I. BACKGROUND 1

II. ACTIVITIES AND RESULTS AT FIFTH GRADE 4

Research Design 4

Results 9

Experimental Treatment 9

Parent Participation 21

Summary 26

III. ACTIVITIES AND RESULTS AT SIXTH GRADE 27

Research Design 27

Results 32

Experimental Treatment 32

Parent Participation 48

Pupil and Teacher Interest 51

Summary 59

NOTES AND REFERENCES 62

APPENDIX A 63

APPENDIX B 64

THIRD YEAR RESULTS
IN THE DENVER-STANFORD PROJECT

by

John L. Hayman, Jr.
James T. Johnson, Jr.

The Denver Public Schools and Stanford University's Institute for Communication Research are currently engaged in a joint research project on the context of instructional television. The purpose of the project is to learn how instructional television can best fit into the total teaching situation. A substantial amount of research has established that television is a very effective teaching medium. Ways of combining it with other educational activities must now be considered, and the Denver-Stanford project is a beginning effort in this direction. Kenneth E. Oberholtzer is principal investigator for the Denver Public Schools and Wilbur Schramm is principal investigator for Stanford University. This is one of a number of project progress reports.

I. BACKGROUND

The Denver-Stanford Project on the Context of Instructional Television has been concerned with teaching Spanish to fifth and sixth grade pupils in the Denver Public Schools, and has used television as the basic instructional medium. The project has involved a search for school and home activities which, as additions to the television instruction, would increase the amount of Spanish learned by pupils.

Research in the project was conducted over three full school years, beginning with 1960-61. This report deals with results obtained during the last of those years, the 1962-63 school year. Other findings, plus an extensive discussion of the project's background, have been treated in previous reports (1, 2, 3, 4), and these items will be repeated here only as they apply to the research in 1962-63.

A knowledge of the general hypotheses to be tested in the project is necessary to understand the third year design. These hypotheses, as originally stated, were:

1. Instructional television of a certain kind, reinforced by certain classroom activities, makes for more learning, ceteris paribus, than without this reinforcement.
2. Instructional television of a certain kind, reinforced by certain home activities, makes for more learning, ceteris paribus, than without this reinforcement.
3. Instructional television of a certain kind, reinforced by certain home and classroom activities, makes for more learning, ceteris paribus, than without these reinforcements or with either one alone.
4. Certain kinds of instructional television, combined with certain kinds of home and classroom reinforcements, make for more learning, ceteris paribus, than other kinds and combinations.

Previous reports have shown hypotheses one, two, and three to be completely verified. The research in 1962-63, therefore, was concerned with hypothesis four -- with the systematic variation of different combinations of home and classroom activities to discover that combination which, when used in conjunction with instructional television, would best meet the objectives of an elementary school foreign language program.

Most of the individual activities investigated in 1962-63 had been tried as single practices before, and in this sense the research was repetitive. The combinations of activities were original in large part, however, and a few practices were entirely new. Thus the research design in 1962-63 was both repetitive and original; it represented a building on previous experience and findings, while, at the same time, it allowed the investigation of new ideas.

The research design and analysis of results were also based, in part, on previous experience. In the project generally, outside influences on a child's performance (that is, independent of the child's own ability, his previous learning, etc.) have fallen into three distinct and mutually exclusive (though interacting) sets of variables; and some attempt has been made to take each of these sets into account. The sets may be termed (1) method variables, (2) teacher variables, and (3) home variables, and are defined as follows:

1. _Method variables_ consist of those practices and activities devised to directly affect learning and produce a specific result. Included are the television presentation; the face-to-face instruction by the classroom teacher; the use of tape recorders, record players, and automated materials; certain specifically designated homework; and the like. These are the variables most easily manipulable and, therefore, most easily evaluated through experimental design.
2. _Teacher variables_ consist of personal characteristics of the classroom teacher, and they affect, among other things, the efficacy of many method variables. Teaching variables include the teacher's interest in the subject matter and in new teaching techniques, the number of courses in foreign language, his professional experience, his ability to achieve rapport with his pupils, etc. Many of these variables are extremely difficult to evaluate, though some of them have been approached through survey techniques.
3. _Home variables_, of great importance because they determine in large part the child's interests and attitudes, are directly related to parental characteristics. Included are the extent of the parents' education, their interest in the specific subject matter at hand, their willingness to work with the child, the standards of performance they establish, etc. Some of these variables have also been approached through survey techniques.

II. ACTIVITIES AND RESULTS AT FIFTH GRADE

A. Research Design

Experimental Procedures. Fifth grade is the first year in which pupils have foreign language instruction in Denver. In accord with FLES (Foreign Language in the Elementary School) recommendations, the first year's instruction is entirely audio-lingual. That is, the first year's instruction is devoted entirely to the development of understanding and speaking skills. Development of the other language skills, reading and writing, begins at sixth grade.

The basic element of the audio-lingual instruction at fifth grade was a 15-minute television program, seen every Monday, Wednesday, and Friday from 12:45 to 1:00 p.m. by each pupil. The television experience was therefore the same for everyone. In fact, this television experience was standard throughout the three years of research; the lessons were developed and recorded on videotape prior to use in 1960-61 and were repeated each year thereafter (5).

A second common element was a 15-minute follow-up to the television lesson. This follow-up was conducted by the classroom teacher through an eclectic approach which utilized structure drill, pattern drill, and a variety of other activities.

The research groups at fifth grade, therefore, differed according to those activities which were added to in-school television viewing and teacher-directed eclectic practice. These additional activities were directed toward evaluating and learning more effective uses of certain electronic aids in the classroom. More specifically, the aids consisted of record players and tape recorders which allowed a child to hear Spanish spoken by a native speaker and, in some cases, record his own voice and hear it played back.

Three groups were set up according to this scheme:

1. No electronic aids. Children in this group had only the common elements of TV viewing and eclectic practice. They served primarily as a control group, or in other words as a base against which to compare the performance of other pupils.
2. Electronic aids without feedback. Children in this group had, in addition to TV viewing and eclectic practice, record players or in a few cases tape recorders. This group did not record their own voices. Rather, they listened to specially-prepared lesson segments.
3. Electronic aids with feedback. Children in this group had, in addition to the common activties, tape recorders which were used to record and play back their own voices. They also made use of special lesson segments.

Electronic aids without feedback, with record players only, were tried in 1961-62; therefore, the first two groups in the 1962-63 design constituted an exact replication. This part of the design was replicated because the earlier results failed to provide clear-cut conclusions. Overall, there was no significant difference in overall performance between those who had record players and those who did not, though some specific classes seemed to profit by using them (2). Further investigation led to the suggestion that teachers had not been sufficiently informed as to how the aids should be used. Therefore, in 1962-63 a set of suggested uses of the aids with each lesson was prepared by language experts and distributed to the appropriate teachers. In addition, these teachers attended special training sessions devoted to the electronic aids.

The feedback condition of the third research group had not been tried previously and was thus original in 1962-63. The idea behind feedback is that a pupil can better understand and correct his mistakes in pronunciation when he hears himself speak, especially if he can compare his own performance to that of an expert. Thus, in addition to the opportunity for varied, directed practice in listening and speaking provided under the no-feedback condition, the pupil has further help in pronunciation.

All practice with electronic aids occurred in the classroom. The most immediate reason for this was that no elementary schools in the Denver system had language laboratories. In any case, however, it would seem desirable to keep the electronic aids in the classroom the first year of instruction because of the closer control allowed the classroom teacher. Fernand Marty has commented as follows on the need for close teacher supervision in the beginning year:

> ...the most favorable situation for improvement in pronunciation is when the teacher, the student, and the machine work jointly. The teacher guides the student, he tells him whether he is getting closer to correct pronunciation . . . . (Without the active help of the teacher) the student is generally unable to determine by himself the causes (of his incorrect pronunciation) (6).

In addition to the above three experimental conditions, the research design provided for parent participation. For the third year, parents were given the opportunity to volunteer to help their children in Spanish.

As before, parent participants were asked to view a repeat of the TV lesson in the evening with their children, to use a special guide book which suggested various home activities, and to use phonograph records which were provided to improve listening and speaking skills. (For a more complete discussion of the home activities, see references 4 and 7.)

Thus each child was classified in the design in two ways -- according to his primary research group and according to whether or not his parents participated. The six possibilities under this design are shown on the following page.

No electronic aids	Electronic aids without feedback	Electronic aids with feedback
No parent help	No parent help	No parent help
No electronic aids	Electronic aids without feedback	Electronic aids with feedback
Parent help	Parent help	Parent help

Research Conditions for Fifth Graders
In 1962-63
Figure 1

Assignment to Procedures. Throughout the project, assignment to primary research group has been by class rather than by individual pupil. Administrative conditions in the separate schools render assignment by individuals practically impossible in a study of this size. The number of classes assigned to each group in 1962-63 was determined by the amount of electronic equipment on hand. Otherwise, assignment was on a random basis. Thus, the number of classes to be assigned tape recorders was determined; and, from a list of all teachers in the project, this number was drawn at random. From those remaining, classes to be assigned record players were determined in like manner. Those still remaining were assigned to the no-electronic-aids condition. Under this scheme, approximately one-half of the fifth grade classes had no electronic aids, approximately one-fourth had electronic aids without feedback, and approximately one-fourth had electronic aids with feedback.

As indicated above, assignment to parent participation groups was not based on probability methods. Rather, it depended entirely on whether or not a child's parents volunteered. The number of such volunteers at fifth grade in 1962-63 was 2874, approximately 45 per cent of the total fifth grade population. This was the largest number of parent volunteers obtained during the project; and, in view of the fact that parent participation was in its third year and had thus lost much of the glitter of something new and exciting, this response was most heartening. Apparently, parent participation of this nature can be used by a school system as an on-going project.

Measurement and Analysis. The pertinent language skills, understanding and speaking ability, were measured at the end of each semester through listening comprehension and speaking tests previously developed by project personnel (8, 9). Also, a listening comprehension pre-test was administered the first week of school. Other information gathered for each pupil in the project included IQ, as measured by the Kuhlman-Anderson Test; the paragraph meaning (PM) score from the Stanford Achievement Test; grade point average (GPA) in academic subjects for the two previous school years; sex; amount of Spanish spoken in the home prior to the beginning of the project; and occupation (according to the Hollingshead scale) of the head-of-the-household.

In addition to the information on pupils, the following data were collected: class size, whether or not the Spanish teacher was also the regular classroom teacher, and the experience and formal preparation (TEP) of the Spanish teacher.

The primary technique used to analyze results was the analysis of covariance, through which the influence of certain measureable variables can be controlled or held constant from group to group. Pre-test, IQ, PM, and GPA were used as control variables in each separate analysis. In addition, the first semester test was used as a control in analyzing second semester results when it was desired to reflect only the amount learned the second semester. The dependent variable in each case was one of the language skills. Certain of the other variables were used as secondary independent or mitigating variables, so that their influence on the dependent variables and their interaction with the primary independent variables of experimental treatment and parent participation were measured.

B. Results

1. Experimental Treatment

Listening Comprehension. Overall, there were no significant differences between the primary research groups on either of the listening comprehension tests. Table 1 shows results of the covariance analysis of the second semester final, with and without the first semester final as a control variable. When the first semester final is used as a control, the results reflect only the amount learned during second semester; otherwise, the amount learned the complete year is reflected.

Table 1

ONE-DEMENSIONAL COVARIANCE ANALYSES
OF THE FIFTH GRADE SECOND SEMESTER LISTENING COMPREHENSION TEST
WITH AND WITHOUT THE FIRST SEMESTER TEST USED AS A CONTROL VARIABLE

With the First Semester Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	875	21, 294.872	
Within Groups	873	21,205.866	24.291
Between Groups	2	89.006	44.503

$$F = \frac{44.503}{24.291} = 1.832$$
$$P > .05$$

Without the First Semester Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	876	37,744.560	
Within Groups	874	37,684.533	43.117
Between Groups	2	60.027	30.013

$$F = \frac{30.013}{43.117} = 0.696$$
$$P > .20$$

Means of test scores and of other control variables used in analyses are given in table 2, and semester test scores adjusted by the covariance technique, are in table 3. (Group variances on each test are given in

Appendix A). Since the covariance analyses indicated that no differences between experimental groups were significant, no statistical comparison of individual group means was made.

Table 2

GROUP MEANS AT FIFTH GRADE
ON THE DEPENDENT AND CONTROL VARIABLES
USED IN LISTENING COMPREHENSION TEST ANALYSES

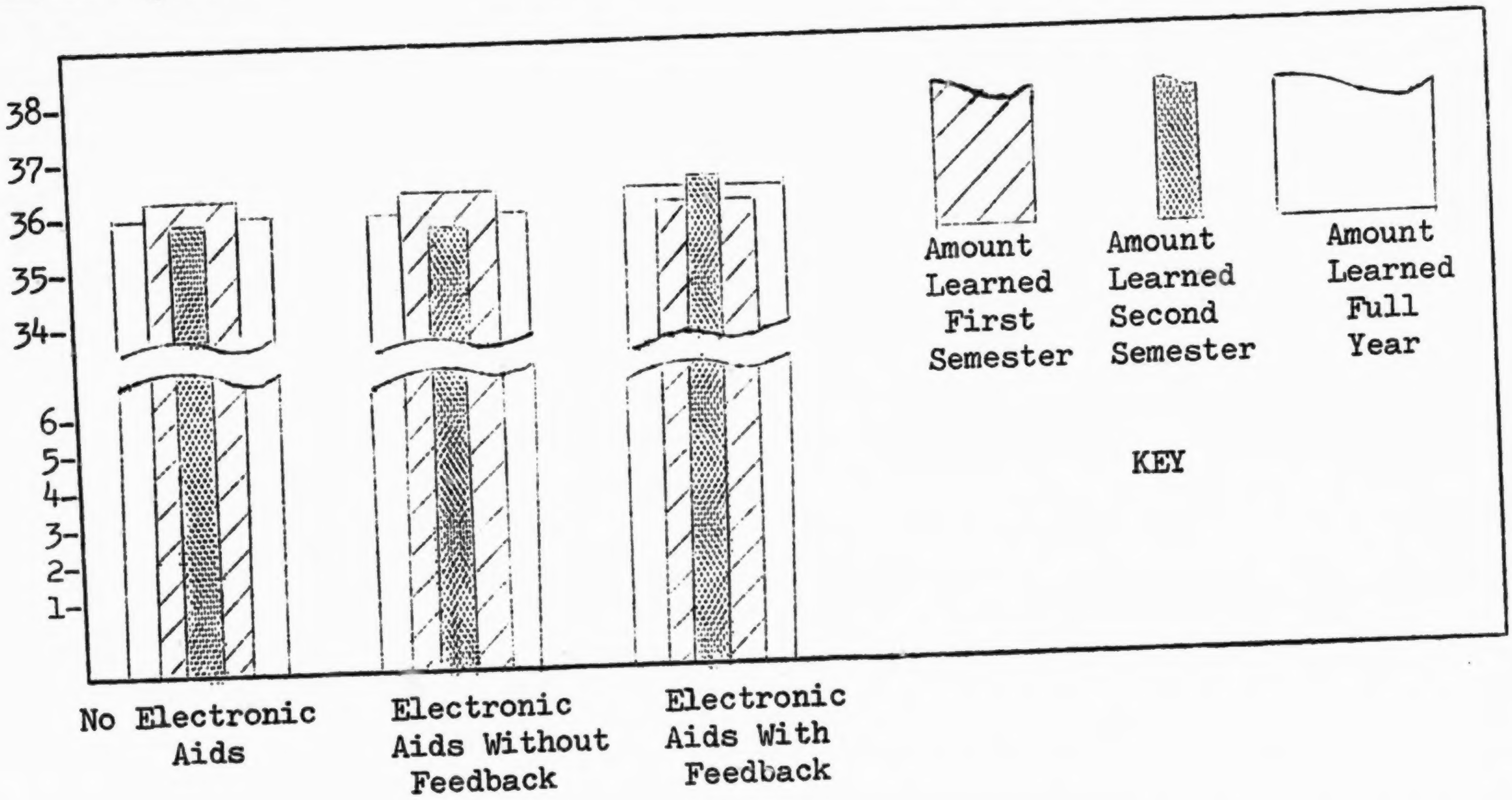
Variable	EXPERIMENTAL GROUP			
	No Electronic Aids	Electronic Aids Without Feedback	Electronic Aids With Feedback	Total
Second Semester Test	35.827	36.279	36.367	36.141
First Semester Test	36.347	36.914	36.173	36.472
Pre-test	8.923	8.664	8.601	8.739
PM	5.528	5.780	5.596	5.630
GPA	2.603	2.711	2.607	2.639
IQ	103.043	105.675	103.831	104.128

Table 3

ADJUSTED GROUP MEANS
ON THE FIFTH GRADE LISTENING COMPREHENSION TESTS

	Experimental Group			
Semester test	No Electronic Aids	Electronic Aids Without Feedback	Electronic Aids With Feedback	Total
First	36.467	36.590	36.359	36.472
Second -- second semester's learning reflected	35.975	35.856	36.620	36.141
Second -- full year's learning reflected	35.988	35.927	36.532	36.141
(N)	(323)	(280)	(278)	(881)

Though no differences were significant, a pattern can be seen in the comparison of means from one semester to the next. This pattern is more evident when the three sets of means in table 3 are plotted on a bar graph, as in figure 2.



Bar Graph of Adjusted Means Which Reflect Amount Learned the First Semester, Second Semester, and Full Year

Figure 2

As table 3 and figure 2 show, all three groups demonstrated quite similar amounts of learning the first semester; the adjusted first semester means were very close. During the second semester, however, the group using electronic aids with feedback achieved more than either of the other two groups, and this results in a higher mean score where learning for the whole year is reflected.

This outcome could be entirely the result of chance since no differences between means were statistically significant. The second semester differences would seem to merit consideration, however, because they fit a pattern observed with an entirely different group of pupils during the 1961-62 school year.

The 1961-62 results indicated that a certain amount of experience and/or preparation is necessary on the part of the classroom teacher before the electronic aids can be used profitably (2, pp. 7-12). The pattern observed in 1962-63 would seem to suggest the same thing. No improvement in listening comprehension skills resulted from the use of electronic aids the first semester, but an improvement may have occurred, at least where feedback was involved, the second semester. A logical explanation for this, in light of past results, would be that a semester's experience was needed before the classroom teachers could use the aids to advantage.

Because of the lack of statistical significance, the observed pattern is at most suggestive, and further information is needed before any more conclusive statements are possible. Such information is, fortunately, available. The experience and preparation of classroom teachers was again measured during 1962-63 so that a further test of the interaction of this teacher variable with the electronic aids method variable is possible.

Results of the two-dimentional analyses, with teacher experience and preparation (TEP) as one dimension and experimental treatment the other, are given in table 4. The first semester listening comprehension test is the dependent variable in one analysis, and the second semester test, with the first semester test as a control variable., is the dependent variable in the other, (As mentioned before, IQ, PM, GPA and pre-test are used as control variable in all such analyses.) Thus, the amount learned the first semester is reflected in one analysis, and the amount learned the second semester is reflected in the other.

Table 4 indicates that there were no statistically significant effects during either the first or second semesters. On the thought that full year results might be different, the second semester test was reanalyzed without the first semester test as a control, but the resulting F-ratios showed very little change. These analyses, therefore, shed no light on the problem.

Table 4

TWO-DIMENSIONAL COVARIANCE ANALYSES
OF FIFTH GRADE LISTENING COMPREHENSION TESTS
EXPERIMENTAL TREATMENT BY TEACHER EXPERIENCE AND PREPARATION

First Semester Test

Source of Variation	Residual			F-ratio	Probability
	Degrees of Freedom	Sum of Squares	Mean Square		
TEP	2	60.391	30.196	.622	>.20
Interaction	4	141.123	35.281	.726	>.20
Within Groups	903	43,836.972	48.546		

Second Semester Test

Source of Variation	Residual			F-ratio	Probability
	Degrees of Freedom	Sum of Squares	Mean Square		
TEP	2	99.169	49.584	1.969	> .05
Interaction	4	152.399	38.099	1.513	> .05
Within Groups	902	22,707.087	25.174		

A glimmer appears, however, when the adjusted means for the full year are examined in table 5. There the aids-with-feedback means are 34.951 for the low TEP group, 35.388 for the middle group, and 37.503 for the high group, and the differences between the middle and high groups and the low and high groups are statistically significant. For the first difference, the critical ratio (D_M/σ_{D_M}) is 2.115/1.127 = 1.877, which has probability of approximately .06; and for the second, the critical ratio is 2.552/1.200 = 2.127, with corresponding probability less than .05. Thus, at least for the aids-with-feedback pupils, a relationship exists between the development of listening comprehension skills and the experience and preparation of the classroom teacher. "Feedback pupils" whose teachers ranked high on TEP had an adjusted full-year score of 37.503, which was more than a full point higher than that of any other group in the two-way matrix. This suggests that the aids can be used effectively.

Table 5

ADJUSTED GROUP MEANS
ON FIFTH GRADE LISTENING COMPREHENSION TESTS
EXPERIMENTAL TREATMENT BY TEACHER EXPERIENCE AND PREPARATION

Experimental Treatment	Teacher Experience and Preparation								
	First Semester Test			Second Semester Test -- First Semester as Control			Second Semester Test No First Semester Control		
	Low	Middle	High	Low	Middle	High	Low	Middle	High
No Electronic Aids	35.736	36.279	35.911	36.262	36.490	35.851	35.851	36.465	35.851
Electronic Aids Without Feedback	36.135	37.132	35.510	35.481	35.716	35.998	35.353	36.297	35.425
Electronic Aids With Feedback	36.106	36.668	37.131	35.099	35.136	36.922	34.951	35.388	37.503
Total	35.978	36.651	36.205	35.659	35.861	36.365	35.420	36.101	36.288

Overall, however, the 1962-63 results, like those of 1961-62 related to electronic aids, are far from clear-cut. Many more questions remain unanswered than answered on the relationship between TEP and use of the aids.

So far only the interaction of teacher experience and preparation with experimental treatment has been considered. Information on several other possible influencing variable was collected; these include sex of pupil, prior Spanish spoken in the home, class size, and regular classroom versus special teacher for Spanish. Results of analyses in which these were used as secondary independent variables are shown in table 6. This table gives F-ratios and probabilities for the secondary categorizations and for the interactions between each secondary variable and experimental treatment. In addition, adjusted means for the secondary categories are shown. The second semester listening comprehension test is the dependent variable, and the first semester test is not used as a control so that learning for the whole year is reflected.

The only variable causing significant differences is prior Spanish in the home. Both the secondary categorization differences and the interaction are significant beyond the .001 level. This interaction merits further consideration and, to allow this, adjusted means for the interior cells of the two-dimensional matrix are shown in table 7.

Table 6

TESTS OF DIFFERENCES AND INTERACTIONS
FOUND IN TWO-DIMENSIONAL ANALYSES
SHOWING FULL YEAR'S LEARNING

Secondary Variable	Category	Test Mean	F-Ratio	Probability
Sex	Male	36.824	< 1.000	> .20
	Female	35.272		
Interaction			2.006	> .05
Prior Spanish in Home	Most of the time	39.433	6.084	< .001
	Part of the time	36.576		
	Very little of the time	35.633		
	None of the time	35.880		
Interaction			4.958	< .001
Class Size	33 or fewer pupils	36.003	< 1.000	> .20
	34 or more pupils	36.439		
Interaction			2.048	> .05
Regular or Other Teacher	Has pupils all day	35.653	2.145	> .05
	Has pupils part of day	36.433		
	Teaches Spanish only	36.244		
Interaction			1.920	> .05

Table 7

ADJUSTED MEANS
ON THE SECOND SEMESTER
LISTENING COMPREHENSION TEST
PRIOR SPANISH BY EXPERIMENTAL GROUP

Spanish Spoken in Home Prior to Beginning Program	Experimental Group		
	No Electronic Aids	Electronic Aids Without Feedback	Electronic Aids With Feedback
Most of the time	41.678	37.655	37.026
Part of the time	36.604	35.777	37.808
Very little of the time	34.253	37.557	35.051
None of the time	35.693	34.751	37.171

The interaction (or inconsistency from group to group) seems to be more logically explained when the means in table 7 are examined by row rather than by column. Children from homes where Spanish is spoken most of the time definitely do better without electronic aids. For the others, however, some form of the electronic aids produces the most learning, though the pattern is not clear by any means. Where Spanish is spoken part or none of the time, those with feedback did best; whereas, the very-little-of-the-time group did better without feedback. The latter group did better with feedback than with no electronic aids at all, however.

Just why the electronic aids are relatively so much less effective among children from homes where Spanish is spoken is uncertain. It has been suggested that these children are confused by differences between the Spanish taught in the school program and that spoken in their homes, and that the electronic aids heighten this confusion. Whatever the cause, however, children from Spanish speaking families comprise only about 7.5 per cent of he school population in Denver and would be a considerably smaller group in most other cities in the United States. Even in Denver, therefore, more than nine out of every ten children could profit from the use of electronic aids in some form, and the feedback condition seems the more desirable.

Speaking. Development of speaking skills under the three experimental conditions at fifth grade was also evaluated through covariance analysis. The relative amounts learned each semester and during the complete year were determined through use of speaking tests administered at the end of each semester, and through the selection of control variables. None of the one-dimensional analyses, comparing the performances of the three basic experimental groups, produced statistically significant results. The analysis reflecting learning for the whole year is shown in table 8; adjusted second semester means, with and without the first semester test used as a control variable, plus group numbers are given in table 9. The numbers here are much smaller than those used in the listening comprehension analyses because of the nature of the speaking tests. To achieve the test reliability needed for valid group comparisons, it was necessary that these tests be administered individually to pupils by project personnel. This process limited the number of pupils who could be tested.

The same pattern concerning relative amounts learned each semester is observed here as for the listening comprehension results, only in this case the reversal is complete. During the first semester, children with no aids performed best and those having aids with feedback were poorest. During the second semester, however, the aids-with-feedback condition produced the most learning, while the no-aids condition produced the least. These two trends balance themselves so that, for the complete year, the means are quite similar.

Table 8

ONE-DIMENSIONAL COVARIANCE ANALYSIS
OF THE FIFTH GRADE SECOND SEMESTER SPEAKING TEST

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	171	1,446.927	
Within Groups	169	1,436.774	8.502
Between Groups	2	10.153	5.077

$$F = \frac{5.077}{8.502} = 0.597$$

$P > .20$

Table 9

ADJUSTED GROUP MEANS
ON THE FIFTH GRADE SPEAKING TESTS

Semester	Experimental Group			
	No Electronic Aids	Electronic Aids Without Feedback	Electronic Aids With Feedback	Total
First	7.826	7.603	7.187	7.577
Second -- Second Semester's Learning Reflected	16.167	16.559	16.967	16.511
Second -- Full year's Learning Reflected	16.357	16.629	16.609	16.511
(N)	(71)	(55)	(49)	(175)

The speaking results, therefore, offer the same suggestion as was noted for listening comprehension--that the electronic aids require a semester's experience before they are generally used profitably.

The same secondary independent variables were used for the two-dimensional covariance analyses of speaking skills as of listening comprehension skills. Therefore, the results on teacher experience and preparation again allow a further test of the conjectured relationship between TEP and use of the electronic aids. The interaction between TEP and experimental treatment was not statistically significant, though the marginal categorizations by TEP produced an F-ratio of 7.087 where second semester results were considered. With n_1 of 2 and n_2 of 121, this F-ratio is significant beyond the .01 level. Adjusted means of the second semester speaking test, with the first semester test used as a control, are shown in table 10. Unfortunately, too few pupils with teachers low in TEP were tested to allow inclusion of the low TEP category in the two-dimensional analysis. This occurred because the samples for testing were drawn before the TEP questionnaires were evaluated, making stratification on TEP impossible.

Table 10

ADJUSTED FIFTH GRADE MEANS
ON THE SECOND SEMESTER SPEAKING TEST
RESEARCH GROUP BY TEACHER EXPERIENCE AND PREPARATION

Teacher Experience and Preparation	Experimental Group			
	No Electronic Aids	Electronic Aids Without Feedback	Electronic Aids With Feedback	Total
High	17.678	16.604	17.283	17.213
Middle	15.372	15.875	16.551	15.791

The results in table 10 do not follow the same pattern in the interior cells of the matrix that was found for listening comprehension. The pattern here, rather, resembles that of 1961-62 in which pupils of teachers in the middle TEP position performed relatively better with electronic aids, while those of teachers in the high position did best with no aids.

None of the other interactions was statistically significant, nor did any other secondary categorizations produce significant differences. Nevertheless, because prior Spanish produced significant results on the listening comprehension test, this variable was examined in further detail. The adjusted second semester means, reflecting the full year's learning, are shown in table 11. The limited numbers forced a combination of the "most" and "part of the time" categories. A comparison of table 11 with table 7 reveals that the pattern of means is precisely the same for speaking as for listening comprehension. That is, pupils from homes where Spanish is spoken do better with no electronic aids, while the others do better with some form of the aids. Furthermore, as on listening comprehension, the feedback condition produces the most favorable results where Spanish was not previously spoken in the home.

Table 11

ADJUSTED MEANS
ON THE SECOND SEMESTER
FIFTH GRADE SPEAKING TEST
PRIOR SPANISH BY EXPERIMENTAL GROUP

Spanish Spoken in Home Prior to Beginning Program	Experimental Group		
	No Electronic Aids	Electronic Aids Without Feedback	Electronic Aids With Feedback
Most or Part of the time	19.673	17.145	16.328
Very little of the time	15.522	16.605	15.729
None of the time	16.241	16.547	17.345

Overall, the speaking results generally support those of listening comprehension. They support fairly well the notion that the electronic aids are used to better effect as the classroom teacher gains experience with them, and they offer solid backing to the finding that electronic aids are most valuable if the pupil is not from a home where Spanish has been spoken previously.

2. Parent Participation

For the third consecutive year, parent participation proved of significant value to most pupils. The results in 1962-63 were, in fact, quite similar in every respect to those of the previous year. Parent participation had greater effect the first semester than it did the second, and a significant interaction between parent participation and prior Spanish in the home was found. The results were similar on both listening comprehension and speaking.

Table 12

TWO-DIMENSIONAL COVARIANCE ANALYSES
OF FIFTH GRADE LISTENING COMPREHENSION TESTS
PARENT PARTICIPATION BY PRIOR SPANISH SPOKEN IN HOME

First Semester Test

Source of Variation	Residual Degrees of Freedom	Residual Sum of Squares	Residual Mean Square	F-ratio	Probability
Parent Participation	1	98.982	98.982	2.635	> .05
Prior Spanish in Home	3	2,253.617	751.206	20.000	< .001
Interaction	3	264.564	88.188	2.347	> .05
Within Groups	808	30,349.201	37.561		

Table 12 (cont'd)

TWO DIMENSIONAL COVARIANCE ANALYSES
OF FIFTH GRADE LISTENING COMPREHENSION TESTS
PARENT PARTICIPATION BY PRIOR SPANISH SPOKEN IN HOME

Second Semester Test

Source of Variation	Residual Degrees of Freedom	Residual Sum of Squares	Residual Mean Square	F-ratio	Probability
Parent Particiaption	1	77.067	77.067	1.902	> .05
Prior Spanish in Home	3	1,940.100	646.700	15.064	< .001
Interaction	3	336.349	112.116	2.768	< .05
Within Groups	808	32,736.301	40.510		

Results of two two-dimensional covariance analyses, with parent participation and prior Spanish in the home as the separate independent variables, are shown in table 12. In one analysis, the first semester listening comprehension test is the dependent variable. In the other, the second semester test is the dependent variable, and is not controlled for first semester results. When only second semester learning is reflected (analysis not shown), none of the effects produce significant results.

Adjusted means for these two-dimensional breakdowns are given in table 13. In addition, this table includes the means which reflect only second semester learning.

This table shows that parent participation is most effective where the least amount of Spanish was spoken in the home prior to enrollment in the program. This has been the finding in previous years as well. Where no Spanish was spoken, the situation of a majority of the children, parent help increases learning to a highly significant degree, and it increases learning, though to a smaller degree, for those in the very-little-of-the-time category. On the other hand, where Spanish is spoken in the home most of the time, parent help is definitely detrimental. The occurrence of this effect at statistically significant levels over a three year span, with entirely different groups of pupils and parents, leaves no doubt that the effect is real. For most children, parent help in Spanish can be distinctly advantageous.

Table 13

ADJUSTED FIRST AND SECOND SEMESTER MEANS
FOR FIFTH GRADE LISTENING COMPREHENSION TESTS
PARENT PARTICIPATION BY PRIOR SPANISH SPOKEN IN HOME

Spanish Spoken in Home Prior to Beginning Program	Test					
	First Semester		Second Semester — Semester Reflected		Second Semester — Full Year Reflected	
	Parent Help	No Parent Help	Parent Help	No Parent Help	Parent Help	No Parent Help
Most of the time	42.462	41.893	36.486	38.099	40.571	41.811
Part of the time	36.474	38.667	36.705	36.385	36.865	37.981
Very little of the time	36.798	36.365	35.857	36.015	36.229	36.105
None of the time	36.004	34.446	36.074	35.351	35.926	34.081

The more disturbing and less understandable aspect of this pattern is the detrimental effect of parent help on those children from homes where Spanish is spoken natively. One explanation previously advanced is that there is a confusion between the "school Spanish" taught in the program and the local version of Spanish used at home. Thus, the parent might attempt to help the child, but his own Spanish is so different that he actually hinders. Another explanation, also previously advanced, is that a higher proportion of parent volunteers from the Spanish speaking group than from other groups may have failed to follow through with the activities. The result in this case could have been a negative attitude toward Spanish on the part of the child.

Fortunately, a partial test of the contending explanations was possible through use of the socio-economic status (SES) information available. As previously stated, the occupation of the head-of-household was determined for each child. Occupation is not a perfect indicator of SES, of course, but it is one of the most valid single indicators which can be readily applied to large groups.

It happens that a large proportion of Spanish speaking adults in the Denver area fall into the lower SES categories. A failure by Spanish speakers to continue the parent-participant activities would be expected to relate to this fact of social class, and such a failure would therefore be typical of other low SES groups. In this event, an interaction would exist between occupation and parent participation.

Table 14

TWO-DIMENSIONAL COVARIANCE ANALYSIS
OF THE SECOND SEMESTER LISTENING COMPREHENSION TEST
PARENT PARTICIPATION BY OCCUPATION OF HEAD-OF-HOUSEHOLD

Source of Variation	Residual			F-ratio	Probability
	Degrees of Freedom	Sum of Squares	Mean Square		
Parent Participation	1	198.313	198.313	4.937	< .05
Occupation	2	11.317	5.659	0.141	> .20
Interaction	2	12.258	6.129	0.153	> .20
Within Groups	583	23,420.228	40.172		

To test this idea, children in the "part" and "most of the time" categories on prior Spanish were eliminated so that the occupation categories would be free of Spanish speakers, and parent participation was run against occupation in a two-dimensional analysis. The results with full year learning reflected are given in table 14.

When the children from homes where Spanish is spoken natively are eliminated, the overall difference between those with and those without parent help becomes statistically significant. Neither the interaction nor the differences produced by the categorization by occupation are significant. Adjusted means relating to table 14 are given in table 15. These means are quite consistent; children with parent help do better than those without such help in every occupation category, and, further, occupation shows no relationship to performance. Evidently, a similar proportion of parent-participants at all SES levels followed through with the activities

Table 15

ADJUSTED SECOND SEMESTER
LISTENING COMPREHENSION TEST MEANS
PARENT PARTICIPATION BY OCCUPATION OF HEAD-OF-HOUSEHOLD

Occupational Level	Parent Participation		
	Parent Help	No Parent Help	Total
High	36.364	34.452	35.394
Middle	35.871	35.053	35.441
Low	35.906	34.799	35.235
Total	35.996	34.863	35.384

The problem, therefore, would appear to lie with Spanish spoken natively in the home and not with a failure by Spanish speaking parents to continue with the activities. This finding enhances the value of the parent help method, for it shows that it will work at all levels of social class.

On the speaking results for the full year, the parent help pupils performed significantly better than those without parent help. The F-ratio on the one-dimensional analysis was 4.267 which, with n_1 of 1 and n_2 of 145, is significant beyond the .05 level. The adjusted means were 17.043 for the parent help group and 15.780 for the others. Thus, parent help made a statistically significant contribution to the speaking skill. The lack of numbers on the speaking test made testing the effects of prior Spanish through a two-way breakdown impossible.

C. Summary

In summary, fifth grade results show that the electronic aids, especially those with feedback, are a desirable addition to the classroom Spanish program. An exception to this occurs among children from homes where Spanish is spoken natively. This exception affects very few children, however; the vast majority can profit from the use of electronic aids.

Parent participation, for the third consecutive year, also proved to be valuable. Again the exception is among children of native speakers.

It was found previously that the eclectic method is best for direct teaching, and results over the total span of the project leave no doubt that well-prepared and experienced teachers improve the quality of instruction. A carefully prepared, on-going inservice training program would therefore seem to be a must for any school system combining television, electronic aids, and other new techniques as they have been combined in the Denver-Stanford project.

The following activities at fifth grade are therefore recommended: basic instruction by television, teacher-directed instruction by the eclectic method, electronic aids with feedback, parent help at home, and an on-going teacher in-service training program.

III. ACTIVITIES AND RESULTS AT SIXTH GRADE

A. Research Design

Experimental Procedures. Reading and writing are introduced at sixth grade in the Denver Public Schools elementary Spanish program, and a major portion of project effort at sixth grade in 1962-63 went into the evaluation of ways to introduce the reading and writing skills. At the same time, the teaching of listening comprehension and speaking skills was continued.

The most interesting innovation introduced in the project with regard to reading and writing has undoubtedly been automated instruction. Automated programs, designed to teach children to read and write the Spanish they have already learned to understand and speak, were developed during the first two years of research (10).

In 1961-62, reading and writing instruction entirely by the automated approach and entirely by the teacher-directed approach were compared. Results of this comparison led to the conclusion that automated instruction is indeed an effective and useful teaching device, but that it does not seem capable of carrying the entire teaching load, at least at the elementary school level. In the report on 1961-62 results, it was hypothesized that automated instruction would be most effective when combined in some way with teacher-directed instruction.

This hypothesis was tested in 1962-63 through various combinations of automated and teacher-directed instruction. In setting up research groups, a base for comparison was first established by repeating the automated only and teacher-directed only conditions from the previous year. The automated-only groups were in turn divided so that original (O) and revised (R) versions of the materials could be used. (The two versions are described in a previous report. See note 3). The original version was used in combinations with teacher-directed instruction.

All direct in-school reading and writing instruction occurred during a 30-minute period from 12:45 to 1:15 each Wednesday. In addition, each sixth grade pupil was given audio-lingual instruction on Tuesdays and Thursdays. The audio-lingual instruction was handled in much the same way as for fifth grade; that is, the children first watched a 15-minute television program and then had 15 minutes of teacher-directed instruction by the eclectic method. This instruction was the same for all sixth graders and was therefore constant in the experimental design.

The design incorporated two additional activities which were introduced in 1962-63. One was a special area of the classroom, termed the "Spanish corner", in which Spanish books and newspaper, record players with Spanish recordings, tape recorders, and assorted cultural artifacts were placed. Each child in classes where the corner was used was allowed approximately 30 minutes a week, on a scheduled basis, in the corner. The child was not directly supervised while in the Spanish corner, but was allowed to pursue any activity he liked.

The second new activity was regularly assigned homework. Some of the children were assigned specific sections of the automated instruction booklets to be completed at home, while others were assigned additional reading in elementary-level Spanish readers. (These readers were also used in the classroom in some instances, and use of them was termed "extended reading.")

The research groups, which were derived by combining the different activities in various ways, are shown in table 16. Note that the first four groups consist of activities involving no direct face-to-face instruction by the classroom teacher, while in the last six, the classroom teacher handles at least part of the instruction directly. The activities listed in table 16 are in addition to the standard audio-lingual instruction, of course. The table also shows abbreviated designations which, to conserve space, will be used in the remainder of this report to indentify the different experimental groups. (Unless otherwise indicated, the original version of the automated materials was used.)

Table 16

ACTIVITIES OF RESEARCH GROUPS
AT SIXTH GRADE DURING THE 1962-63 SCHOOL YEAR

Group Activities	Designation
Automated instruction in school	As
Automated instruction in school, revised version of automated materials	As(R)
Extended reading in school and automated instruction at home	EsAh
Extended reading in school and automated instruction at home, revised version of automated materials	EsAh(R)
Teacher-directed instruction	Td
Teacher-directed and automated instruction in school	TdAs
Teacher-directed instruction in school and automated instruction at home	TdAh
Teacher-directed and automated instruction in school and extended reading at home	TdAsEh
Teacher-directed and automated instruction in school and Spanish corner	TdAsC
Teacher-directed instruction in school, automated instruction at home, and Spanish corner	TdAhC

In still another innovation at sixth grade, seven teachers were set apart from the research design and given complete freedom to handle the classroom instruction in any way they liked. (Their pupils viewed the TV lessons.) The only restrictions were that each teacher in this group was required to keep a complete, written record of his activities relative to Spanish instruction, and each was asked to attend special meetings at which the different activities were discussed. These teachers were designated the "tomorrow's classroom" group.

Finally, parent participation was encouraged at sixth grade as well as at fifth grade. This was the second year in which parent-help was tried at sixth grade, and its use was essentially a replication of the 1961-62 trial which showed that parent help continues to be beneficial to pupils in their second year of Spanish instruction. As at fifth grade, pupils with parent help were scattered throughout the research design so that each research group consisted of pupils with and without parent help.

Assignment to Procedures. Random assignment was employed at sixth grade so far as possible, though conditions necessitated some departure from it. In the first place, some classroom teachers handled both fifth and sixth grade Spanish instruction. Since these teachers were occupied with fifth grade audio-lingual instruction during the 12:45 to 1:15 period each Wednesday, they were not available for reading and writing instruction. Teachers in this situation were assigned to the two automated-only conditions, As and As(R), which required no direct teaching on Wednesdays. These teachers, therefore, provided the audio-lingual instruction for their sixth grade pupils on Tuesdays and Thursdays, and they were replaced on Wednesdays by other teachers who acted mainly as proctors.

The remaining teachers, except for the seven in the "tomorrow's classroom" group, were randomly assigned to the eight other research groups. The seven with unrestricted classroom practice were selected in two stages. First, the proposed treatment was explained to all of the sixth grade teachers, and volunteers were requested. Some 50 teachers responded. Then the seven were chosen randomly from those who volunteered. This self-selection procedure was followed deliberately to assure interested teachers in the "tomorrow's classroom" group.

As in all previous cases, parent participation was on a volunteer basis. The only restriction at sixth grade was that parents must also have participated the previous year when their child was in fifth grade. Of some 3,000 parents who participated in 1961-62, 1770 elected to continue their participation in 1962-63. A substantial (though undetermined) number of the remainder did not continue because of unavoidable circumstances, such as moving from the city.

Measurement and Analysis. The two dependent variables at fifth grade, listening comprehension and speaking, served the same role at sixth and were measured through special instruments which covered sixth grade content. In addition, reading and writing was of concern at sixth grade and served as a third dependent variable. This skill was measured at the end of each semester through 66-item reading and writing tests.

The control and secondary independent variable were about the same at sixth grade as for fifth. IQ, PM, GPA, and listening comprehension pre-test were used as standard control variables, and sex, occupation of the head-of-household, class size, and teacher experience and preparation (TEP) were used as secondary independent variables.

An addition was teacher and pupil attitude toward Spanish as an academic subject. Attitude was measured through a set of scales devised especially for the project by Paul I. Jacobs and Milton H. Maier of the Educational Testing Service. Three separate scales measured pupil pre-attitude during the first week of the school year, teacher attitude about two-thirds of the way through the first semester, and pupil post-attitude at the end of the first semester. The reported reliabilities for these scales were approximately .50 for pupil pre-attitude, .85 for teacher attitude, and .86 for pupil post-attitude. A complete description of these scales and their development is available from the Educational Testing Service (11).

Attitude results were used almost entirely as secondary independent variables, though, as will be seen in the results section, pupil post-attitude was used as a dependent variable in one or two cases.

As stated previously, the differences between research groups at sixth grade depended almost entirely on the type of reading and writing instruction involved, and the reading and writing skill would therefore seem to be the most important dependent (or criterion) variable. This is not true, however, because the language skills interact. The instructional procedure which is best in terms of reading and writing is not necessarily the best in terms of listening comprehension or speaking, though ideally this would be the case. All skills were analyzed, therefore, for a complete evaluation of research procedures. Covariance analysis was again the primary statistical technique. As at fifth grade, IQ, PM, GPA, and listening comprehension pre-test were used as controls in all analyses which employed one of the language skills as dependent variable. Also, first semester results were used as control variables when it was desired to reflect only the amount learned the second semester.

B. Results

1. Experimental Treatment

Listening Comprehension. One-dimensional analyses of the second semester test were made with and without the first semester test used as a control variable, and results of these are shown in table 17. Highly significant F-ratios occurred in both cases. Adjusted means are shown in table 18, and they prove to be quite similar, indicating that the different treatments had relatively the same effect throughout the school year.

Table 17

ONE-DIMENSIONAL COVARIANCE ANALYSES
OF THE SIXTH GRADE SECOND SEMESTER LISTENING COMPREHENSION TEST
WITH AND WITHOUT THE FIRST SEMESTER TEST USED AS A CONTROL VARIABLE

With the First Semester Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	1,848	51,990.781	
Within Groups	1,839	51,175.995	27.828
Between Groups	9	814.786	90.532

$$F = \frac{90.532}{27.828} = 3.253$$

$P < .001$

Without the First Semester Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	1,849	69,324.729	
Within Groups	1,840	68,079.610	37.000
Between Groups	9	1,245.119	138.347

$$F = \frac{138.347}{37.000} = 3.739$$

$P < .001$

Table 18

ADJUSTED SIXTH GRADE MEANS
ON THE SECOND SEMESTER LISTENING COMPREHENSION TEST

Experimental Group	Test Mean		(N)
	Second Semester's Learning Reflected	Full Year's Learning Reflected	
As	31.873	31.168	(164)
As(R)	32.281	32.312	(179)
EsAh	33.050	33.086	(112)
EsAh(R)	31.967	31.929	(117)
Td	32.618	32.978	(222)
TdAs	32.952	33.066	(201)
TdAh	32.033	32.309	(230)
TdAsEh	31.508	31.022	(174)
TdAsC	33.769	33.735	(242)
TdAhC	32.235	32.407	(213)

Table 18 shows the prediction regarding the combination of automated and teacher-directed instruction to be at least partially confirmed. The TdAs condition is clearly superior to either As or As(R), and it is slightly better than Td, though the difference is not statistically significant. However, by adding one more ingredient, the Spanish corner, to the combination a dramatic increase in learning occurs. The best performance by far was achieved by those children in the TdAsC condition, that is, by those with automated and teacher-directed instruction in school and with a Spanish corner. In terms of the listening comprehension (or understanding) skill, therefore, TdAsC is the desirable combination.

The two-dimensional breakdown, with experimental groups divided according to teacher experience and preparation (TEP), produces results which support quite strongly some previous conclusions regarding the teacher variables. These results were similar each semester, so only the analysis reflecting the full year's learning is shown. Table 19 gives this analysis, and table 20 shows the corresponding adjusted group means.

Table 19

TWO-DIMENSIONAL COVARIANCE ANALYSIS OF THE
SECOND SEMESTER SIXTH GRADE LISTENING COMPREHENSION TEST
EXPERIMENTAL TREATMENT BY TEACHER EXPERIENCE AND PREPARATION

Source of Variation	Residual			F-ratio	Probability
	Degrees of Freedom	Sum of Squares	Mean Square		
TEP	1	43.358	43.358	1.253	> .20
Interaction	9	1,775.604	197.289	5.449	< .001
Within Groups	1,830	66,258.648	36.207		

The means for those six groups with teacher-directed instruction form a confusing pattern and suggest no ready conclusions. Among the four groups without teacher-directed instruction in reading and writing, however, the pattern is much more consistent and meaningful. In three of these groups, pupils whose teachers rated high on the TEP scale performed better than those whose teachers rated low, and the difference between means in the one exception, AS(R), was quite small. Furthermore, the differences between the low and high TEP pupils in the EsAh and EsAh(R) groups were by far the largest produced by the break on TEP. Thus the conditions with the least amount of direct supervision in reading and writing, where pupils were more on their own, were the ones on which the experience and preparation of the teacher handling the audio-lingual instruction had the most effect. This supports the conclusion, stated in a previous report (2), that the teacher exerts a motivating effect that is most noticeable when the pupil is free to do more or less as he likes. A more stringent test of this hypothesis will occur when reading and writing skill rather than listening comprehension is the dependent variable.

Two-dimensional analyses were also run with class size and sex as the secondary independent variables, and the only statistically significant effect was an interaction between class size and experimental treatment. No such analysis was possible with the regular-versus-special teacher variable because of an unfortunate distribution of numbers; only two of the ten research groups had at least two teachers in every cell of the matrix. A one-dimensional analysis showed no significant difference between pupils of special teachers and those of regular teachers.

The F-ratio for the interaction between class size and experimental treatment was 3.462, which with n_1 of 9 and n_2 of 1,757, is significant beyond the .001 level. The adjusted group means for the two-dimensional breakdown are given in table 21.

Table 20

ADJUSTED MEANS AT SIXTH GRADE ON THE SECOND SEMESTER LISTENING COMPREHENSION TEST EXPERIMENTAL TREATMENT BY TEACHER EXPERIENCE AND PREPARATION

Experimenatl Treatment	Teacher Experience and Preparation	
	Low	High
As	30.699	31.885
As(R)	32.434	32.157
EsAh	32.489	36.201
EsAh(R)	30.561	36.721
Td	32.931	33.092
TdAs	33.563	31.965
TdAh	32.016	33.086
TdAsEh	31.168	30.708
TdAsC	32.979	35.268
TdAhC	33.451	30.834
Total	32.351	32.722

Table 21

ADJUSTED SECOND SEMESTER LISTENING COMPREHENSION TEST MEANS EXPERIMENTAL TREATMENT BY CLASS SIZE

Experimental Treatment	Class Size	
	29 or fewer	30 or more
As	29.731	32.602
As(R)	30.722	32.282
EsAh	34.921	32.441
EsAh(R)	31.137	32.269
Td	33.667	32.803
TdAs	33.941	33.175
TdAh	33.413	31.938
TdAsEh	30.082	31.582
TdAsC	35.085	33.205
TdAhC	33.040	31.987
Total	32.472	32.454

The means in table 21 fit a meaningful pattern when considered by experimental treatment. In five of the six treatments involving teacher-directed instruction, pupils in smaller classes learned the most; whereas, in three of the four treatments with no teacher-directed instruction, the higher means were achieved by those in larger classes. Thus, smaller classes seem advantageous only where direct teacher instruction is involved. Once more, however, the analysis of reading and writing results will provide a more stringent test.

To summarize, the listening comprehension results suggest that a combination of teacher-directed and automated reading and writing instruction in school, plus a Spanish corner, is desirable. Furthermore, they suggest that a direct relationship exists between amount learned and the experience and preparation of the classroom teacher and that this relationship is strongest when the pupil is on his own. Finally, smaller classes seem important only for direct, face-to-face teaching.

Reading and Writing. One-dimensional covariance analyses of the second semester reading and writing test are given in table 22 and the related adjusted means are in table 23. Reading and writing results, like listening comprehension, were quite similar both semesters, so only second semester and full year learning is shown in the tables.

The strong relationship between the reading and writing and listening comprehension skills is evident in table 23, for the reading and writing results are quite similar to those of listening comprehension discussed previously. As before, teacher-directed instruction alone is better than automated instruction alone, but the combinition of the two produces still more learning. In this case, however, the difference between the Td and TdAs conditions is larger than it was for listening comprehension -- in fact, the difference of 2.543 between the means of these groups where the full year is reflected is itself significant beyond the .05 level. So the hypothesis that the combination would be superior to either practice alone is definitely confirmed.

Table 22

ONE-DIMENSIONAL COVARIANCE ANALYSES OF THE SIXTH GRADE SECOND SEMESTER READING AND WRITING TEST WITH AND WITHOUT THE FIRST SEMESTER TEST USED AS A CONTROL VARIABLE

With the First Semester Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	1,848	79,600.470	
Within Groups	1,839	78,230.491	42.540
Between Groups	9	1,369.979	152.220

$F = \frac{152.220}{42.540} = 3.578$

$P < .001$

Without the First Semester Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	1,849	131,295.900	
Within Groups	1,840	126,666.912	68.841
Between Groups	9	4,628.998	514.332

$F = \frac{514.332}{68.841} = 7.471$

$P < .001$

Table 23

ADJUSTED SIXTH GRADE MEANS ON THE SECOND SEMESTER READING AND WRITING TEST

Experimental Group	Test Mean		(N)
	Second Semester's Learning Reflected	Full Year's Learning Reflected	
As	15.924	13.970	(164)
As(R)	16.885	16.166	(179)
EsAh	16.907	16.785	(112)
EsAh(R)	15.569	15.188	(117)
Td	17.346	17.474	(222)
TdAs	18.778	20.017	(201)
TdAh	17.254	17.711	(230)
TdAsEh	17.710	17.367	(174)
TdAsC	17.622	18.608	(242)
TdAhC	18.313	18.056	(213)

Results for both listening comprehension and reading and writing skills, therefore, indicate that a combination of activities is desirable. There is some disagreement on which combination is best, however. The listening comprehension results showed the addition of a Spanish corner to the teacher-directed and automated in school combination to be desirable, but table 23 shows less learning for this condition than for TdAs alone. This disagreement is not of major proportion, however, since the TdAsC means are quite close to those of TdAs. The larger difference -- where full year learning is reflected -- is 1.409 points, which is not statistically significant. A look back at table 18 shows small differences between these groups on listening comprehension also, so either combination would seem desirable in terms of the skills so far examined.

At least two other noteable points appear in table 23. First, automated instruction appears to be more useful as school work than as homework. In the two cases where As and Ah can be compared -- TdAs versus TdAh and TdAsC versus TdAhC -- the As means are generally higher. (Table 18 shows the As means are also higher on listening comprehension results.) Note, however, that in both comparisons Td is a common element, so that the classroom teacher could have been assisting the children with their automated instruction in school. Whether a pure As-Ah comparison would give different results cannot be determined from the results at hand. It would make little difference if a change did occur, however, because the desirability of teacher-directed instruction as a part of the combination has been firmly established.

The second point involves the original and revised versions of the automated materials. The 1961-62 results indicated rather firmly the desirability of the original version (3). Table 23 does not give such a clear-cut picture, however. For in-school use, the revised version seems better. Furthermore, it is not clear that the difference in school and home environments is a contributing factor here because the home-use comparisons are contaminated by another variable -- extended reading in school. It is thus not possible to state with any confidence which version of the automated materials seems best.

The two-way breakdown of reading and writing results by experimental treatment and by teacher experience and preparation confirms even more strongly than the listening comprehension results the previous conclusions regarding TEP. The two-dimensional covariance analysis is given in table 24 and the adjusted means in table 25.

Table 24

TWO-DIMENSIONAL COVARIANCE ANALYSIS OF THE
SECOND SEMESTER SIXTH GRADE READING AND WRITING TEST
EXPERIMENTAL TREATMENT BY TEACHER EXPERIENCE AND PREPARATION

Source of Variation	Residual				
	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio	Probability
TEP	1	1,038.749	1,038.749	15.555	< .001
Interaction	9	3,424.632	380.515	5.698	< .001
Within Groups	1,830	122,203.531	66.778		

Table 25

ADJUSTED MEANS AT SIXTH GRADE
ON THE SECOND SEMESTER READING AND WRITING TEST
EXPERIMENTAL TREATMENT BY TEACHER EXPERIENCE AND PREPARATION

Experimental Treatment	Teacher Experience and Preparation	
	Low	High
As	14.142	13.711
As(R)	16.068	16.296
EsAh	15.303	24.535
EsAh(R)	13.419	21.378
Td	17.245	18.048
TdAs	20.793	18.277
TdAh	17.594	18.018
TdAsEh	16.623	18.979
TdAsC	17.013	21.840
TdAhC	17.377	19.079
Total	16.848	18.449

Both the difference between means of the low and high TEP categories and the interaction are highly significant statistically. As the "total" figures indicate, there is a direct realtionship between TEP and test score. This relationship is generally demonstrated through the table, in fact, since the high TEP group had the highest mean score in eight of the ten research procedures. Overall, however, the differences tend to be relatively small, on the order of one to two points.

In view of this general pattern, the differences for the two groups with the least amount of teacher supervision -- EsAh and EsAh(R) -- are rather astounding. In each group, children with high TEP teachers have means about 60 per cent higher than those with low TEP teachers. This finding is even more remarkable when it is remembered that the teacher of reference here is the one who handles audio-lingual instruction on Tuesdays and Thursdays and has little to do with the pupil's reading and writing activities. That the characteristics of this teacher have such an impact on the acquistion of reading and writing skills -- much more impact than on listening comprehension skills with which the teacher is directly involved -- suggests strongly that the teacher plays a major role in motivating the pupil. In other words, the teacher inspires the child to work in those situations where he is not under direct supervision. This type of influence might well be felt under many different conditions. It could carry over into homework, for example. The value of homework per se in any subject might depend more on the quality of instruction provided in class than on any other single factor. Table 25 shows that means of the three additional groups involving homework -- TdAh, TdAsEh, and TdAhC -- do, in fact, support this notion.

In any event, the importance of teacher experience and preparation is established, and the statement made previously, that an ongoing inservice program is vital in a program of this nature, has received additional support.

Regarding the other secondary independent variables, the situation was much the same for reading and writing as for listening comprehension. The analysis matching experimental treatment against class size was the only one which produced statistically significant results. The derivation of F-ratios for this analysis is shown in table 26, and the adjusted means are in table 27.

The directions of differences in table 27 follows those of the listening comprehension results (table 21) exactly. In three of the four groups with no teacher-directed instruction, children in larger classes scored higher on the reading and writing test; and, in five of the six groups with teacher-directed instruction, performance was higher where the number in the class was smaller. This outcome again suggests that having smaller classes is important only in the direct teaching situation where individual attention is required.

An even stronger test of the class size effect would be provided by comparing very large classes -- with, perhaps, 40 pupils or more -- to very small classes -- with 20 pupils or less. There were too few such extreme cases to allow a valid comparison, however. The distribution of classes by class size is given in Appendix B.

Table 26

TWO-DIMENSIONAL COVARIANCE ANALYSIS OF THE
SECOND SEMESTER SIXTH GRADE
READING AND WRITING TEST
GROUP BY CLASS SIZE

Source of Variation	Residual			F-ratio	Probability
	Degrees of Freedom	Sum of Squares	Mean Square		
Class Size	1	318.691	318.691	4.642	< .05
Interaction	9	2,648.756	294.306	4.287	< .001
Within Groups	1,757	120,612.319	68.674		

Table 27

ADJUSTED SECOND SEMESTER
READING AND WRITING TEST MEANS
EXPERIMENTAL TREATMENT BY CLASS SIZE

Experimental Treatment	Class Size 29 or fewer	 30 or more
As	13.388	15.340
As(R)	15.125	16.276
EsAh	22.167	14.767
EsAh(R)	14.504	15.948
Td	19.091	16.986
TdAs	20.389	19.952
TdAh	18.933	17.270
TdAsEh	16.389	18.112
TdAsC	21.690	17.374
TdAhC	18.692	17.595
Total	17.871	16.967

This raises interesting team teaching possibilities. Large classes could be used for viewing the television lessons and for working with automated materials, and small classes could be used for teacher-directed instruction related to the eclectic audio-lingual method, the reading and writing activities, and the use of electronic aids. With careful scheduling in this kind of arrangement, each child could have more individual attention with no increase in the number of teachers.

Speaking. The one-dimensional analysis of the second semester speaking test, with full year's learning reflected, is given in table 28. The numbers are small, of course, because of the nature of the test.

The adjusted means are given in table 29. Because of the small numbers, only five of the groups are represented in this table, but these form a meaningful pattern.

Table 28

ONE-DIMENSIONAL COVARIANCE ANALYSIS
OF THE SIXTH GRADE SECOND SEMESTER SPEAKING TEST

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	214	5,893.861	
Within Groups	209	5,333.580	25.520
Between Groups	5	560.281	112.056

$$F = \frac{112.056}{25.520} = 4.391$$

$$P < .001$$

Table 29

ADJUSTED MEANS ON THE
SECOND SEMESTER SPEAKING TEST

Experimental Treatment	Test Mean
As(R)	21.290
Td	22.825
TdAs	20.510
TdAsC	20.226
TdAhC	24.787
Total	21.688

The two highest means in table 29 are for those practices where the classroom teacher spent the full 30 minutes each Wednesday in direct, face-to-face instruction. This is understandable. Assuming that speaking is improved only through actual practice, a child who spends part of the available time with automated instruction or with reading only is not likely to develop his speaking skills as highly as one who speaks Spanish the whole time -- even though the speaking takes place in what is primarily reading and writing instruction. Thus, if speaking is the criterion, the combination of automated and teacher-directed instruction in school would be less desirable than teacher-directed instruction alone.

Which procedure, then, is most effective for sixth grade pupils? This depends on which skill is to be emphasized since the best results on each skill were produced by different procedures. The fact is, of course, that all of the skills are important and should be emphasized, and a compromise solution is needed.

Fortunately, a very good compromise seems available. The results on listening comprehension and reading and writing indicated, as previously noted, that either the TdAs (teacher directed and automated instruction in school) or TdAsC (teacher-directed and automated instruction in school and a Spanish corner) conditions would be desirable. The results in table 29 suggest that the Spanish corner may be of some aid in developing the speaking skills, presumably because of the electronic aids in the corner. The TdAsC condition is not significantly different from TdAs, but TdAhC is significantly better than Td. Though the results are not clear-cut, the Spanish corner shows potential as an aid to speaking.

The compromise, therefore would be TdAsC. With this procedure, conditions are close to optimum for both the listening comprehension and reading and writing skills, and the potential for improving speaking skills is present. Since this potential presumably exists because of the presence of electronic aids in the Spanish corner, the situation is similar to that at fifth grade. Teachers need special preparation in directing use of the aids, and they need experience. Therefore, the effectiveness of the TdAsC procedure would be dependent upon an accompanying inservice training program.

2. The "Tomorrow's Classroom" Treatment

As noted previously, seven sixth grade teachers were removed from the experimental design and allowed to teach Spanish entirely as they wished. These teachers were selected randomly from a group who volunteered for this method, so a selection factor was operating. Self-selection was deliberately allowed in this instance in an attempt to get highly interested teachers in the group.

The results obtained by these teachers are quite exciting. One-dimensional covariance analyses comparing results of the "tomorrow's classroom" group with those of all others combined, are given in table 30, and adjusted group means are in table 31. The dependent variables here are the second semester listening comprehension and reading and writing tests, and learning for the full year is reflected.

Pupils in the "tomorrow's classroom" group did considerably better than the others on both tests, and the difference on reading and writing is little short of astounding. The "tomorrow's classroom" mean on reading and writing is almost 50 per cent higher than the other mean. These results show what highly motivated and highly interested teachers can do for pupils.

This outcome should not be taken to mean it is best to let every teacher do what he wishes, however, in spite of the fact that the "tomorrow's classroom" teachers had a free hand. Quite the contrary, these teachers used all of the procedures and aids available to the regular research groups, and they invented a number of new ones on their own. The results show, rather, the value of the materials and methods now at hand when used by a really interested teacher, and they suggest that even more varied activities are useful.

Table 30

COVARIANCE ANALYSES
OF SECOND SEMESTER TESTS
"TOMORROW'S CLASSROOM" COMPARED TO OTHER GROUPS

Listening Comprehension Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	2,115	80,300.023	
Within Groups	2,114	79,131.008	37.432
Between Groups	1	1,169.015	1,169.015

$$F = \frac{1,169.015}{37.432} = 31.230$$

$p < .001$

Reading and Writing Test

Source of Variation	Residual		
	Degrees of Freedom	Sum of Squares	Mean Square
Total	2,115	156,920.219	
Within Groups	2,114	150,751.932	71.311
Between Groups	1	6,168.287	6,168.287

$$F = \frac{6,168.287}{71.311} = 86.498$$

$p < .001$

Table 31

ADJUSTED MEANS
ON THE SECOND SEMESTER TESTS
"TOMORROW'S CLASSROOM" COMPARED TO OTHER GROUPS

Test	Mean	
	"Tomorrow's Classroom"	Experimental Groups Combined
Listening Comprehension	35.932	32.574
Reading and Writing	25.220	17.527

Several of the "tomorrow's classroom" teachers came up with interesting activities not otherwise tried in the project, and a number of these should be considered in designing an elementary school Spanish program. A description of the "tomorrow's classroom" teachers and the methods they employed will be provided in a subsequent report.

3. Parent Participation

Parent participation proved valuable again at sixth grade, and, in fact, the decline in its effect the second semester, noted over three years at fifth grade, did not occur. Covariance analyses of listening comprehension results for the two semesters are shown in table 32, and the adjusted means are in table 33.

The lack of a second semester decline, though heartening, is a little puzzling since it has been noted consistently in other results. Perhaps it is related to the double selection involved. Parent participation at sixth grade was restricted to those parents who had participated the previous year when their children were fifth graders, and not all of the parents elected to continue the second year. It would seem safe to assume that, although some drop-outs occurred because of moving and other uncontrollable reasons, many parents did not continue because of low interest in the method. Those who did continue would tend to have high interest. Selection was involved first at fifth grade and then, after a year's trial, at sixth grade, and those parents still in the program would be those most likely to follow through with the assigned activities.

Whether the double selection explanation is valid or not, these listening comprehension results give a very strong support to parent particiaption as an aid to pupil learning.

On tests of the speaking skill, parent participation made very little difference. None of the F-ratios were significant; in fact, they were all less than 1.00. And, as would be expected from these results, the differences between means were quite small. The adjusted means, showing learning for each semester and the full year, are given in Table 34.

Table 32

ONE-DIMENSIONAL COVARIANCE ANALYSES OF THE SIXTH GRADE LISTENING COMPREHENSION TESTS PARENT PARTICIPATION VERSUS NO PARENT PARTICIPATION

First Semester Test

Source of Variation	Residual Degrees of Freedom	Residual Sum of Squares	Residual Mean Square
Total	1,874	68,251.012	
Within Groups	1,873	68,060.093	36.337
Between Groups	1	190.919	190.919

$$F = \frac{190.919}{36.337} = 5.254$$

$P < .05$

Second Semester Test

Source of Variation	Residual Degrees of Freedom	Residual Sum of Squares	Residual Mean Square
Total	1,873	52,843.630	
Within Groups	1,872	52,623.202	28.111
Between Groups	1	220.328	220.328

$$F = \frac{220.328}{28.111} = 7.838$$

$P < .01$

Table 33

ADJUSTED MEANS ON THE SIXTH GRADE LISTENING COMPREHENSION TESTS PARENT PARTICIPATION VERSUS NO PARENT PARTICIPATION

Test	Means: No Parent Participation	Means: Parent Participation
First Semester	31.079	32.572
Second Semester	32.257	33.011

Table 34

ADJUSTED MEANS ON THE SIXTH GRADE SPEAKING TESTS PARENT PARTICIPATION VERSUS NO PARENT PARTICIPATION

Semester Test	Means: No Parent Participation	Means: Parent Participation
First	15.722	15.874
Second -- Second Semester's Learning Reflected	22.278	22.027
Second -- Full Year's Learning Reflected	22.248	22.126

The activities suggested for participants in the parent-help method have been entirely audio-lingual in nature, that is, they have involved listening to and speaking Spanish but have not involved reading and writing. On the surface, it would seem that parent help should influence only the listening comprehension and speaking skills. However, if the prior conclusion that the method is partly effective because of increased motivation on the part of the child to learn Spanish is valid, then pupils with parent help should perform better than the others on reading and writing tests. Pupils with parent help did perform better than those without such help on each reading and writing test, though the differences were not statistically significant for either semester alone. Where full year learning was reflected, however, the F-ratio was 5.428, and for n_1 of 1 and n_2 of 1,490, this F-ratio has probability less than .05. The means for these analyses are shown in table 35.

Table 35

ADJUSTED MEANS ON THE
SIXTH GRADE READING AND WRITING TESTS
PARENT PARTICIPATION VERSUS NO PARENT PARTICIPATION

Semester Test	Means	
	No Parent Participation	Parent Participation
First	22.180	22.945
Second -- Second Semester's Learning Reflected	17.079	17.636
Second -- Full Year's Learning Reflected	16.967	18.024

These results on reading and writing appear to validate the hypothesis that at least part of the effectiveness of parent participation is motivational. Not only does the child perform better on tests of the understanding skill, he also demonstrates statistically significant superiority on reading and writing.

Overall, the sixth grade results constitute another strong endorsement of the parent-participant method. Two of the language skills, listening comprehension and reading and writing, are definitely improved through parent help, and on speaking, the third skill, pupils whose parents participated did at least as well as those whose parents were not involved.

4. Pupil and Teacher Interest

Language Skills as Dependent Variables. As discussed previously, a set of teacher and pupil interest scales was developed by Educational Testing Service especially for the project and was used at sixth grade during 1962-63. These scales were validated in terms of internal consistency only and not against any external criterion. Therefore, results derived from their use must be interpreted with some caution, though their usefulness is not necessarily impaired.

Assuming that the scales actually measured teacher and pupil interest in Spanish, the first question considered was, "What effect does teacher and pupil interest have on the amount pupils learn?" To answer this first question, the interest scale results were analyzed entirely as independent variables, with language skills as dependent variables. (First semester tests were used here because the interest scales were administered during the first semester.) Table 36 gives the results of two two-dimensional analyses in which teacher interest and pupil post interest were the independent variables. Listening comprehension is the dependent variable in one analysis, and reading and writing is dependent in the other.

The analyses show that both teacher interest and pupil interest produced significant differences on tests of the language skills and that there was no interaction. Furthermore, the larger effect was produced by pupil interest; the

F-ratios for this variable are on the order of 3 1/2 times larger than those for teacher interest. The adjusted means in table 37 were generally in the direction expected, that is, higher test means are related to higher interest.

Table 36

ANALYSES OF THE EFFECTS OF TEACHER AND PUPIL INTEREST ON SIXTH GRADE LANGUAGE SKILLS TESTS

Listening Comprehension Test

Source of Variation	Residual Degrees of Freedom	Residual Sum of Squares	Residual Mean Square	F-ratio	Probability
Teacher Interest	2	343.459	171.730	4.694	< .01
Pupil Interest	2	1,044.801	522.401	14.277	< .001
Interaction	4	62.978	15.744	0.430	> .20
Within Groups	1,703	62,308.597	36.588		

Reading and Writing Test

Source of Variation	Residual Degrees of Freedom	Residual Sum of Squares	Residual Mean Square	F-ratio	Probability
Teacher Interest	2	1,107.863	553.932	5.617	< .01
Pupil Interest	2	4,121.816	2,060.908	20.897	< .001
Interaction	4	202.430	50.608	0.513	> .20
Within Groups	1,703	167,950.138	98.620		

Table 37

ADJUSTED MEANS
ON THE LANGUAGE SKILLS TESTS
TEACHER INTEREST BY PUPIL INTEREST

Pupil Interest	Listening Comprehension Test				Reading and Writing Test			
	Teacher Interest				Teacher Interest			
	Low	Middle	High	Total	Low	Middle	High	Total
High	32.969	32.761	33.175	32.946	23.533	22.427	25.627	23.733
Middle	32.213	32.071	33.416	32.509	24.432	23.599	24.708	24.165
Low	30.754	30.360	31.567	30.816	20.904	19.607	21.051	20.376
Total	31.947	31.713	32.795	32.101	23.087	22.032	23.881	22.868

Since both teacher interest and pupil interest produce statistically significant differences in amount learned, it will be well to examine each effect in more detail. First, with regard to teacher interest, the question arises as to whether interest operates independently from experience and preparation. One would suspect that, in general, the teachers higher in experience and preparation would be more likely to be higher in interest. In this case, the interest scales could be merely remeasuring teacher experience and preparation (TEP). To determine the relationship, a two-way table, with interest by TEP, was set up for sixth grade teachers, and related chi square computed. The results are shown in table 38.

Table 38 indicates that, contrary to what might be logically expected, there was no correlation between teacher interest and teacher experience and preparation. The contingency coefficient for the X^2 in table 38 is .13.

Both teacher variables are important, therefore. Both have been shown to relate significantly to the amount pupils learn, and the results here show that they are independent. In this case, the best performance should be by pupils with teachers high in experience and preparation and also high in interest. This notion was tested through a two dimensional covariance analysis with TEP and interest as independent variables and the understanding and reading and writing skills as dependent variables. As in previous analyses, the independent variables produced statistically significant differences. Also significant, however, were the interactions. With listening comprehension as the dependent variable, the interaction F-ratio was 5.954, and with reading and writing as dependent, it was 2.334. The related probabilities, for n_1 of 4 and n_2 of 1.823, were, respectively, less than .001 and approximately .06.

Table 38

RELATIONSHIP BETWEEN TEACHER INTEREST AND TEACHER EXPERIENCE AND PREPARATION

Teacher Interest	Teacher Experience and Preparation			Significance Test
	Low	Middle	High	
High	27.59%	23.88%	30.00%	
Middle	51.72	40.30	41.67	$X^2 = 2.734$
Low	20.69	35.82	28.33	$.50 < P < .70$
Total	100.00%	100.00%	100.00%	
(N)	(29)	(67)	(60)	

The best performances on both listening comprehension and reading and writing tests were, as hypothesized, by pupils with teachers high on each independent variable. To this extent, the results are not surprising. The interaction, however, came as a complete surprise, and the pattern associated with it suggests some revealing conclusions regarding the teacher variables.

The high teacher interest and high experience and preparation condition is clearly desirable. But what happens when teacher interest is low? In that situation, experience and preparation does not have the advantageous effects one would expect. Indeed, for the listening comprehension skill, experience and preparation is a detriment if interest is low! The poorest performance of all was by pupils of teachers high in experience and preparation but low in interest. One might conclude, on observing these results, that a teacher can be expert in not teaching, as well as the other way around.

Table 39

ADJUSTED MEANS
ON THE LANGUAGE SKILLS TESTS
TEACHER INTEREST BY EXPERIENCE AND PREPARATION

Teacher Interest	Listening Comprehension Test			Reading and Writing Test		
	TEP			TEP		
	Low	Middle	High	Low	Middle	High
High	31.725	32.243	33.586	21.237	22.398	23.967
Middle	30.610	32.564	31.187	21.390	22.713	21.335
Low	32.922	32.227	31.015	20.841	23.768	22.961

Which of the teacher variables under consideration is more important? Both have been shown to relate significantly to pupil learning. Furthermore, they have been found independent -- a teacher's experience and preparation has nothing to do with his interest. Yet they interact, so that high levels of both are needed. This being the case, it is difficult to say which is more important. The fact remains, however, that experience and preparation can be dealt with by a school system through careful recruitment, inservice training, and summer workshops, while interest would seem to be much more of an individual matter with less chance for outside control.

Pupil Interest as Dependent Variable. Tables 36 and 37 showed pupil interest to exert even more influence on learning than teacher interest. Therefore, pupil interest also merits close examination, though, within the scope of the Denver-Stanford project, little more can be done with it as an independent variable. More important in this context is to determine what will effect pupil interest, that is, to treat pupil interest as a dependent rather than an independent variable.

Accordingly, pupil interest (measured at the end of the first semester) was treated as a dependent variable in several analyses. In these, an attempt was made to determine the effect on pupil interest of the following factors: teacher interest, teacher experience and preparation, parent participation, experimental treatment, and pupil interest prior to beginning the sixth grade program (called "pupil pre-interest"). Significance tests, showing the influence of each of these on pupil interest, are given in table 40.

Table 40

INFLUENCE OF
SEVERAL FACTORS
ON PUPIL INTEREST

Effect	F-ratio	Degrees of Freedom		Probability
		n_1	n_2	
Teacher Interest	0.683	2	1,750	$> .20$
TEP	1.468	2	1,750	$> .05$
Parent Participation	1.614	1	1,829	$> .05$
Experimental Treatment	4.923	9	1,308	$< .001$
Pupil Pre-interest	162.485	3	1,702	$< .001$

Of the five factors, only experimental treatment and pupil pre-interest had statistically significant effects on pupil interest, and a comparison of F-ratios shows that the most influence by far was exerted by pupil pre-interest. In other words, a pupil's interest in Spanish after a semester's work was determined largely by what he thought of it before he began the semester. Adjusted means, with pupil interest categorized by pre-interest and teacher interest, are shown in table 41. The interaction for this analysis was not statistically significant.

Table 41

ADJUSTED
PUPIL INTEREST MEANS
PUPIL PRE-INTEREST BY TEACHER INTEREST

Pupil Pre-interest	Teacher Interest Low	Middle	High	Total
High	14.013	14.331	14.860	14.415
High Middle	12.651	12.477	12.475	12.524
Low Middle	9.048	9.488	10.025	9.529
Low	6.590	7.829	7.450	7.369
Total	10.941	11.213	11.362	11.186

The great influence of pre-interest on pupil interest, and the much smaller influence of teacher interest, are clearly shown in table 41. The means in every category of teacher interest are ranked precisely by pupil pre-interest, while teacher interest has little effect anywhere.

These results have an obvious implication -- that what happens to a child before he starts sixth grade Spanish is vitally important in shaping his attitude toward the subject in sixth grade. Sixth grade, of course, is the second year of Spanish for the child, and it follows a full year of work with a fifth grade teacher. Further, it is the second year in which the child's parents will or will not have participated. So it is not surprising that his attitudes toward Spanish are pretty well set when he starts sixth grade.

While these results pinpoint the importance of the pupil's fifth grade experience, however, they do not relieve the sixth grade teacher of responsibility. This point is emphasized by table 40. Though the largest influence by far is pupil pre-interest, experimental treatment at sixth grade also produced differences significant beyond the .001 level. Table 42 gives adjusted means when pupil interest is categorized by teacher interest and by experimental treatment. The interaction for this breakdown was significant beyond the .05 level.

Table 42

ADJUSTED
PUPIL INTEREST MEANS
EXPERIMENTAL TREATMENT BY TEACHER INTEREST

Experimental Treatment	Teacher Interest			
	Low	Middle	High	Total
As	10.706	10.567	10.908	10.703
As (R)	10.340	9.807	10.761	10.217
Td	10.178	9.118	11.904	10.364
TdAs	9.504	10.288	9.078	9.884
- - - -	- - - -	- - - -	- - - -	- - - -
TdAh	12.053	12.043	12.187	12.087
TdAsEh	11.372	13.862	9.875	12.106
TdAsC	11.700	12.588	11.353	11.932
TdAhC	11.905	11.581	11.344	11.607
EsAh	10.780	11.497	12.129	11.393
EsAh (R)	10.744	12.289	12.837	12.285

Most noticeable in table 42 is a clear division point, marked by the dashed line, between relatively low and relatively high interest. The means below this line are consistently higher than those above it. What separates those practices below the line from those above it? The answer seemed to be that those practices below the line involve a greater variety of activities. All practices above the line had all instruction in the classroom during the regular 30 minute period each Tuesday, Wednesday, and Thursday. All six practices below the line had some activity outside of this regualr period, and five of the six involve homework. The implication is clear: One way to increase a child's interest in Spanish at sixth grade is to add some outside activity, and homework appears to be as effective an additional activity as any.

In a final analysis, teacher interest was the dependent variable and regular versus special teacher was the independent variable. The reader should recall that "regular versus special" refers to the teacher of subjects other than Spanish. A "regular" teacher was one who had the pupils for Spanish as well as all other academic subjects, and a "special" teacher was one who taught them Spanish only. The F-ratio for this analysis was 6.119, which, with n_1 of 1 and n_2 of 175, is significant beyond the .02 level. The mean interest scores were 23.058 for special teachers and 21.316 for regular teacher; therefore, the special teachers showed significantly more interest in Spanish than did the regular teachers.

Insofar as teacher interest *per se* is concerned, the special teachers seem to have an advantage,and this apparently counterbalances other advantages which accrue to regular teachers since results at both fifth and sixth grade showed no significant difference in amounts learned by pupils with special teachers and those with regular teachers. These results are important in their implication to an elementary school Spanish program. At present, many of the regular classroom teachers are not prepared to teach Spanish so that a special teacher must be used. The results show that the special teachers do the job quite well and that a child taking Spanish from a special teacher is likely to learn as much as one with a regular teacher.

C. Summary

In summary, the sixth grade results confirmed the hypothesis that a combination of automated and teacher-directed instruction would be superior to either method alone. This combination produced significantly superior results on both the listening comprehension and reading and writing tests when both types of instruction were given in the classroom in a complimentary manner.

This combination produced somewhat less favorable results on speaking test, apparently because there were fewer chances to pratice speaking when automated instruction was used. However, it appears this deficiency can be partly overcome through use of electronic aids as an additional in-school activity which is scheduled outside of the regular Spanish instruction period. In this instance, the aids, together with additional reading material and cultural artifacts, were used as part of a Spanish corner.

Results derived from special teacher and pupil interest scales showed that these variables indeed affect the amount a pupil learns. Teacher experience and preparation also affects learning, but not always in direct fashion. Rather, teacher experience and preparationinteracts with teacher interest so that larger amounts of experience and preparation are valuable only when interest is high.

Pupil interest related, as a dependent variable, to two factors. The stronger was pupil interest prior to the beginning of the school year, emphasizing the importance of the pupil's experience in fifth grade. The second factor was experimental treatment. Pupil interest was significantly higher under those conditions which provided some activity in addition to those of the regularly scheduled classroom period. Homework was one such additional activity.

Parent participation, for the second year, was a valuable addition to the sixth grade activities. In fact, the interaction between prior Spanish and parent participation was not found at sixth grade; the method was effective for all pupils.

Finally, pupils learned about the same amount whether they were taught by their regular classroom teacher or by a special teacher who handled only the Spanish instruction.

The following activities at sixth grade are therefore recommended: basic audio-lingual instruction by television, teacher-directed audio-lingual instruction by the eclectic method, reading and writing instruction in class by a combination of the teacher-directed and automated methods, a Spanish corner in school, parent help at home, some additional homework, and an on-going inservice teacher training program.

Notes and References

(1) Hayman, John L., Jr. and James T. Johnson, Jr. "Results of the First Year's Research in the Denver-Stanford Project." Denver-Stanford Project on the Context of Instructional Television. Report number 5. Denver, Colorado: Title VII Office, January, 1962. (Mimeo.)

(2) Hayman, John L., Jr. and James T. Johnson, Jr. "Audio-Lingual Results in the Second Year of Research -- 1961-62." Denver-Stanford Project on the Context of Instructional Television. Report number 8. Denver, Colorado: Title VII Office, June, 1963. (Mimeo.)

(3) Hayman, John L., Jr. and James T. Johnson, Jr. "Reading and Writing Results in the Second Year of Research -- 1961-62." Denver-Stanford Project on the Context of Instructional Television. Report number 7. Denver, Colorado: Title VII Office, May, 1963. (Mimeo.)

(4) Mayers, Alan, John L. Hayman, Jr., and James T. Johnson, Jr. "Causative Factors and Learning Related to Parent Participation." Denver-Stanford Project on the Context of Instructional Television. Report number 11. Denver, Colorado: Title VII Office, in process.

(5) The complete set of videotapes developed for use in the project are now available on a rental basis through the Great Plains Regional Television Library, University of Nebraska, Lincoln, Nebraska.

(6) Marty, Fernand L. *Language Laboratory Learning*. Wellesley, Massachusetts: Audio-Visual Publications, 1960. p. 18.

(7) Hayman, John L., Jr. and James T. Johnson, Jr. "A New Opportunity for Parent Participation in Education." Denver-Stanford Project on the Context of Instructional Television. Report number 1. Denver, Colorado: Title VII Office, April, 1961. (Mimeo.)

(8) Andrade, Manual, John L. Hayman, Jr., and James T. Johnson, Jr. "Measurement of Listening Comprehension via Television in Elementary School Spanish Instruction." Denver-Stanford Project on the Context of Instructional Television. Report number 3. Denver, Colorado: Title VII Office, October, 1961. (Mimeo.)

(9) Andrade, Manual, John L. Hayman, Jr., and James T. Johnson Jr. "Measurement of Speaking Skills in Elementary Level Spanish Instruction." Denver-Stanford Project on the Context of Instructional Television. Report number 9. Denver, Colorado: Title VII Office, July, 1963. (Mimeo.)

(10) Barcus, Delbert, John L. Hayman, Jr., and James T. Johnson Jr. "Development of Programed Learning Materials for Use with Televised Spanish Instruction." Denver-Stanford Project on the Context of Instructional Television. Report number 4. Denver, Colorado: Title VII Office, November, 1961. (Mimeo.)

(11) Maier, Milton H. and Paul I. Jacobs. "Evaluating a Program: Some Recommendations and Results." A paper presented at the 48th annual convention of the National Association of Secondary Schools Principals. Princeton, New Jersey: Educational Testing Service, March, 1964. (Mimeo.)

APPENDIX A

Standard Deviations and Variances on Each Test -- By Research Group

Fifth Grade

Research Group		Listening Comprehension First Semester	Listening Comprehension Second Semester	Speaking First Semester	Speaking Second Semester
No Electronic Aids	σ	9.042	8.177	2.460	4.544
	σ^2	81.758	66.863	6.052	20.648
Electronic Aids Without Feedback	σ	7.131	7.482	2.460	4.224
	σ^2	50.851	55.980	5.560	17.842
Electronic Aids With Feedback	σ	7.376	7.938	2.367	4.139
	σ^2	54.405	63.012	5.603	17.131
Total	σ	7.985	7.886	2.429	4.366
	σ^2	63.769	62.197	5.902	19.070

Sixth Grade

Research Group		Listening Comprehension First Semester	Listening Comprehension Second Semester	Reading and Writing First Semester	Reading and Writing Second Semester	Speaking First Semester	Speaking Second Semester
As	σ	6.819	7.021	10.612	10.417		
	σ^2	46.499	49.294	112.615	108.514		
As(R)	σ	8.606	8.577	13.307	12.030	5.485	6.306
	σ^2	74.063	73.565	177.076	144.721	30.085	39.766
EsAh	σ	7.464	7.441	10.490	12.181	4.954	8.325
	σ^2	55.711	55.368	110.040	148.377	24.542	69.306
EsAh(R)	σ	8.833	8.946	11.082	10.753	5.326	6.809
	σ^2	78.022	80.031	122.811	115.627	28.366	46.362
Td	σ	9.477	8.299	12.490	11.831		
	σ^2	89.814	68.873	156.000	139.973		
TdAs	σ	8.160	7.953	12.700	12.245		
	σ^2	66.586	63.250	161.290	149.940		
TdAh	σ	8.086	8.768	13.587	12.235	4.511	6.725
	σ^2	65.383	76.878	184.607	149.695	20.349	45.226
TdAsEh	σ	8.882	8.725	12.962	12.329	4.910	4.610
	σ^2	78.890	76.126	168.013	152.004	24.108	57.912
TdAsC	σ	7.355	7.587	11.819	11.847		
	σ^2	54.096	57.563	139.689	140.351		
TdAhC	σ	8.973	8.703	11.491	10.763		
	σ^2	80.515	75.742	132.043	115.842		
Total	σ	8.342	8.308	12.396	11.857	5.060	7.068
	σ^2	69.599	69.038	153.675	140.608	25.608	49.965

APPENDIX B

Distribution of Classes by Class Size

Fifth Grade

Number of Pupils	Number of Classes	Percent
0 - 19	3	1.6%
20 - 24	11	5.8
25 - 29	34	18.1
30 - 34	70	37.2
35 - 39	50	26.6
40 - 44	15	8.0
45 - 49	3	1.6
50 or more	2	1.1
Total	188	100.0%

Sixth Grade

Number of Pupils	Number of Classes	Percent
0 - 19	8	4.4%
20 - 24	14	7.8
25 - 29	49	27.2
30 - 34	58	32.2
35 - 39	37	20.6
40 - 44	9	5.0
45 - 49	3	1.7
50 or more	2	1.1
Total	180	100.0%